# How Patient is Venture Capital?*

Namrata Narain
Harvard University

Please click here for the most recent version.

September 15, 2024

**Abstract**

I document that venture capital is disproportionately allocated to technologies with short lags between investment and returns; this allocation resembles that by public firms, and is shorter-term relative to the distribution of commercial value and scientific opportunity. To make these comparisons, I propose an empirical measure of the lag between investment and returns across technologies and, using natural language processing, assign innovation funding over the past four decades from venture capital firms, federal agencies, and public firms to each technology. The prevailing 10-year, closed-ended fund structure in the industry is partly responsible for the concentration of funding in technologies with short lags. I find that fund managers invest in short-term innovation when nearing a fund's term limit, which is consistent with term limits affecting investment decisions. Two instruments designed to isolate variation in the supply of available funds confirm these results.

*Email:* nnarains@gmail.com. I am grateful to Gabe Chodorow-Reich, Josh Lerner, and Adi Sunderam for their advice and support. This paper also benefited from comments by Maxim Alekseev, Abhijit Banerjee, Shai Bernstein, Adrien Bilal, Kevin Chen, Leonardo D'Amico, Ed Glaeser, Claudia Goldin, Justin Katz, Larry Katz, Vincent Rollet, Kunal Sangani, Peleg Samuels, Sankalp Singayapally, Lawrence Summers, Lingxuan Wu, and seminar participants at the Harvard University Macroeconomics and Finance Workshops.

# 1. Introduction

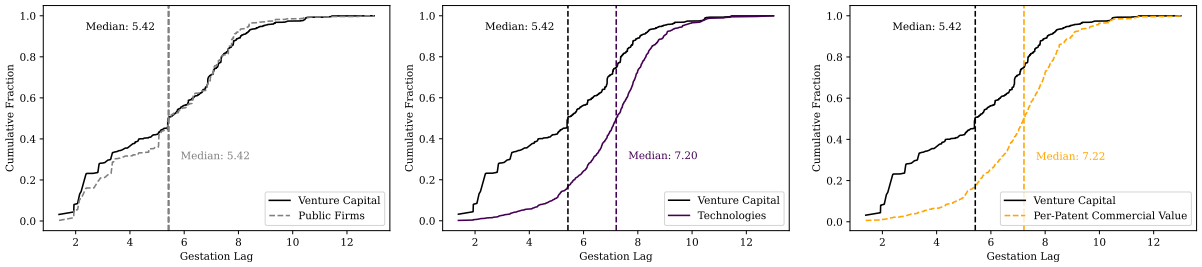
Financing innovation is a hard problem; it requires upfront investment and delivers uncertain returns at unpredictable horizons (Arrow, 1962). Several studies have documented that investor pressure can lead public firms to prioritize short-term cash flows and reduce investment in innovation (Summers and Balls, 2015; Asker et al., 2015; Bernstein, 2015; Terry, 2023). Venture capital is thought to mitigate such pressures through its organizational structure. For instance, venture capital firms invest through closed-ended funds, which prevent investors from withdrawing their money on demand, and managers' compensation is linked to returns from investments rather than short-term metrics such as quarterly earnings reports that plague public firm managers (Lerner and Nanda, 2020; Terry, 2023). However, there is little evidence if venture capital fulfills its promise of being a patient source of finance for long-term innovation.

In this paper, I evaluate the allocation of venture capital to long-term innovation and determine whether its allocation resembles the distribution of commercial value and scientific opportunity in the economy.[1] I present my key finding in Figure 1: the allocation of venture capital across innovations with different lags between investment and returns, i.e. gestation lags, is identical to the allocation by public firms. Furthermore, venture capital is disproportionately allocated to short-term innovation relative to the full set of technologies in the economy and the distribution of commercial value across these technologies. The remainder of this paper shows that venture capital is also short-term relative to two benchmark government programs that finance small firms seeking to commercialize or adopt new technology: the Small Business Innovation Research (SBIR) and Small Business Technology Transfer programs (STTR). Additionally, venture capital is short-term relative to the distribution of four federal agencies that finance innovation in the U.S. economy: National Science Foundation, the Department of Defense, the National Institutes of Health, and the Department of Energy.

To evaluate the allocation of venture capital, I construct a novel dataset that makes three contributions: it determines a set of technologies that captures economy-wide innovation since 1980, develops an empirical measure of gestation lags for these technologies,

[1] Various definitions, often interchangeable, have been proposed for innovation, invention, and technology in the economics literature (Bryan and Williams, 2021). I define "technology" as any method of combining inputs to produce goods and services currently consumed, and "innovation" as improvements in these methods. I do not use the term "invention," which often refers to new goods or services produced following innovation. These definitions align with Ruttan (1959).

**Figure 1:** Distribution of Venture Capital




**(a)** Public Firms **(b)** All Technologies **(c)** Commercial Value

*Notes:* Each figure shows the cumulative distribution of venture capital funding to technologies with gestation lags measured on the x-axis. The solid black line represents venture capital investments made between 1980 and 2022, with all amounts converted to 2022 constant dollars. The dashed grey line in Figure 1a presents the cumulative distribution of global research and development expenditure of public firms. The purple line in Figure 1b shows the cumulative distribution of gestation lags in the U.S. economy. The orange dashed line in Figure 1c presents the cumulative distribution of patent-level commercial value across technologies as measured by (Kogan et al., 2017). Vertical lines in each figure correspond to the median value for each cumulative distribution. For further information on the construction of the underlying data, see Section 3.

and assigns funding from major innovation funding sources to these technologies. As my set of technologies, I use the nearly 600 four-digit codes from Combined Patent Classification (CPC) system that is used to classify patents, and which provides a granular description of these technologies. The CPC system enables me to construct a proxy for gestation lags by analyzing the patents assigned to each code. To assign funding to these technologies, I compile descriptions of technologies, ventures, public firms, and federal research awards. Using natural language processing, I match each venture, firm, and federal award to the technology to which its description is most similar.

Why does the allocation of venture capital firms look similar to that of public firms given the objective of the former to be a source of patient capital? In the second part of the paper, I provide one explanation. Venture capital firms form limited partnerships—"funds"— with investors, typically structured with a 10-year limit and extendable by one to two years. This contract structure has been a mainstay of the industry since its inception and is also prevalent across other private capital sources such as buyout private equity, infrastructure, real estate, and natural resources (Berlin, 2014; Lerner and Nanda, 2020; Brown et al., 2021).

This contract design affects the investment of venture capital into long-term innovation for two reasons. First, managers choose to invest in innovation whose value is realized within the term limit. Limited Partnership Agreements (LPAs) prevent managers

from investing in ventures they had previous interests in, thus preventing managers from rolling over investments across funds. Additionally, severe informational asymmetry regarding the value of the venture discourages sales of unrealized ventures to managers in other firms. Second, managers prefer investments they can exit quickly because successful exits assist in raising subsequent funds.

A key challenge in estimating the effect of the industry-wide adoption of 10-year funds on its investment in long-term innovation is that there is no variation in term limits across venture capital firms. I address this challenge by comparing the investment decisions of fund managers at different stages of the fund; within-fund comparisons approximate the experience of investors facing different term lengths (Barrot, 2017). I show that fund managers invest in younger ventures earlier in the fund's life, and they exit from these ventures sooner relative to those invested in later stages. Both findings reject the null hypothesis that term limits do not matter for investment decisions.

The relationship between term limits and investment in young ventures also holds in the aggregate. I find a strong negative correlation between the availability of young funds in a year and the age of ventures that receive a first-time investment in that year. However, this could reflect a mechanical relationship if more venture capital funds are raised at the same time as more new ventures enter the economy. I instrument for the availability of young funds with exogenous increases in supply of capital to venture capital firms. Large institutions that are Limited Partners in venture capital funds face restrictions on the share of their portfolio that they can invest in risky assets such as venture capital (Brown et al., 2021). Therefore, when returns increase elsewhere in the economy, such as public stock markets, Limited Partners are able to increase their investment in venture capital. Using 5-year returns in the S&P 500 as my instrument, I find that younger ventures receive investment when more young funds are available in the economy due to higher public market returns.

A second instrument I develop exploits differences across venture capital firms in their exposure to Limited Partners with varying levels of market sensitivity. I focus on the period in the aftermath of the dot com bubble (2000-2007), when funding for the venture capital asset class declined sharply (Townsend, 2015). I show that venture capital firms backed by endowments before 1999 raised more new funds between 2000 and 2007 and also invested in younger ventures in this time relative to firms backed by pension funds. The different market sensitivities of endowments and pension funds have been discussed previously, and I document them here using contemporaneous returns in the venture

capital sector (Swensen, 2000; Ivashina and Lerner, 2019).

The empirical evidence on the effect of term limits on mangers' investment decisions presents a theoretical puzzle: the finite term limit should not affect the managers' investment decisions, even if LPAs prevent rolling over investments across funds. At any time in the fund's life, managers should be able to exit investments, recover their capital, and be compensated for any knowledge gained about those investments. In a later draft, I show that term limits affect investment decisions only when managers expect to face discounts on the secondary market, and that secondary market discounts arise endogenously in this setting.

Finally, I address the efficiency of the near-universal adoption of the same structure across venture capital funds. The uniformity could be either be efficient or reflect inefficiency, with venture capital firms incapable of deviating from the standard structure. My model outlines a set of conditions under which it is efficient for venture capital firms and Limited Partners to not deviate from a default term length. The legal literature offers support for the alternative hypothesis. Studies document the wide-spread use of contract defaults in settings beyond venture capital and argue that this "stickiness" could arise due to a third actor–the law firm–that determines contract terms between parties. I review this literature, which documents and explores the consequences of the extensive use of legal boilerplate templates (Gulati and Scott, 2011; Richman, 2011; Choi et al., 2017; Nyarko, 2021), in a later section.[2]

**Contributions to the Literature.** My paper contributes to two strands of literature: (1) documenting frictions in the private sector when financing innovation, and (2) understanding the drivers of the aggregate direction of innovation.

Recent studies document substantial returns from investment in innovation, ranging from 50% to 300%, suggesting significant under-funding of innovation (Jones and Summers, 2021; Fieldhouse and Mertens, 2023; Moretti et al., 2023). Banerjee and Duflo (2014); Buera et al. (2015) survey numerous studies that document the financial frictions facing entrepreneurs. Beyond overall under-investment in innovation, there is specific evidence that the private sector under-invests in long-term innovation (Budish et al., 2015). To address these gaps, governments finance innovation themselves, and a large body of literature evaluates the impact of government funding on innovation (Howell, 2017; Babina

[2]In contrast to venture capital, other sources of innovation funding, such as the federal government, use a menu of different contracts. In Appendix D, I provide a brief overview of the different types of contracts used by federal agencies to finance innovation.

et al., 2023; Myers and Lanahan, 2022; Dyevre, 2023). A key finding in this literature is that governments tend to target innovations with greater spillovers; this finding has been used to explain the dual trends of declining productivity growth and declining share of federal investment in research and development in the U.S. economy. My paper provides a novel method to assign innovation funding from different sources to the same set of technologies and compare their allocations directly. Furthermore, this analysis proposes a concrete approach to identifying the innovations that are underfunded by the private sector and could therefore benefit from government support—specifically, those with longer lags between investment and returns.

Patents are a key policy tool that governments use to encourage private sector investment in innovation. Recent studies find that while patents encourage innovation, their fixed 20-year duration may limit investment in long-term innovation (Acemoglu, 2011; Budish et al., 2015; Sampat and Williams, 2019; Moscona, 2021). The 10-year terms that I study impose a shorter timeline of investment than patent length. Moreover, since the term limits are set by voluntarily adopted contracts, changes to patent lengths may be meaningless to increase investment in long-term innovation. My paper establishes a novel channel to explain private sector under-financing of innovation, which has implications for the optimal design of innovation policy (Bloom et al., 2019; Kremer et al., 2021).[3]

Second, this paper contributes to the literature on the determinants of the direction of innovation. Studies such as Arrow (1962); Akcigit and Kerr (2018); Bryan and Lemus (2017); Hopenhayn and Squintani (2021); Acemoglu et al. (2022); Moscona and Sastry (2022) explore various axes of heterogeneity in innovations—incremental vs. radical, basic vs. applied, etc.—and examine how the market allocates capital across these types of innovations. Understanding how competitive forces shape innovation will be crucial in addressing future challenges like climate change (Acemoglu et al., 2012).[4] To this literature, I add an empirical investigation of one dimension of heterogeneity in innovation:

[3]Several studies document corporate short-termism, for instance Graham et al. (2005). Gompers (1996) shows that career concerns amongst venture capital fund managers can push ventures to go public prematurely. Terry (2023) highlights how short-term pressures in public firms, arising due to quarterly earnings calls, distort investment in research and development. Surveys of this literature can be found in Asker et al. (2015) and Stein (2003). The theoretical counterpart of this literature explores how asymmetric information between financial intermediaries and investors affects investment decisions (Diamond, 1989). Optimal contracts, discussed in (Mayer, 2022; Thomas, 2019; Garfagnini and Strulovici, 2016; Manso, 2011; Admati and Pfleiderer, 1994; Stein, 2005; Arrow, 1993), can help mitigate some of these frictions.

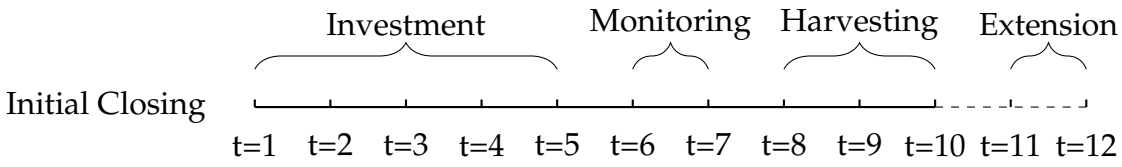
[4]Another branch of research studies the choice of innovation by scientists themselves, for example Azoulay et al. (2011); Myers and Tham (2023).

gestation lags. I also highlight a new mechanism—voluntarily adopted contracts—due to which the venture capital industry may prefer to finance technologies that have short gestation lags.[5]

## 2. Design of the Venture Capital Fund

> "Our industry is still beholden to a rigid 10-year fund cycle pioneered in the 1970s. As chips shrank and software flew to the cloud, venture capital kept operating on the business equivalent of floppy disks." (Roelof Botha, Sequoia Capital, 2021)

**Figure 2:** The Lifecycle of a Venture Capital Fund




Venture capital firms organize themselves as Limited Partnerships, which allow two types of investors with distinct roles: the General Partners ("fund managers") have broad discretion regarding the ventures they fund, while Limited Partners ("savers") provide the capital (Kaplan and Stromberg, 2003; Ivashina and Lerner, 2019). I present a schematic of the lifecycle of a typical venture capital fund in Figure 2. The term limit of 10 years begins at the fund's initial closing, which is the first date at which the General Partners deem that enough capital has been committed to meet the fund's objectives. Fund managers have significant discretion regarding which ventures to invest in, when to provide follow-on funding, and when to exit and return capital to the Limited Partners. However, LPAs often restrict investments into new ventures to the first 3-5 years following the initial closing, known as the investment period. After investment, fund managers may make sequential follow-on investments in some ventures during the monitoring period, after which they exit and harvest their returns. Contracts can usually be extended for

[5]Other studies that have attempted to empirically measure heterogeneous gestation lags are Hall et al. (1986); Sternitzke (2010); Budish et al. (2015); Mansfield (1991, 1995, 1998); Toole (2012). I review the literature on gestation lags in Appendix B.1. As such, my paper is one of the first to empirically study the allocation of private capital across technologies with heterogeneous gestation lags.

one to two years; any further extensions require approval from a majority of the Limited Partners or their advisory committee.[6]

LPAs also impose meaningful restrictions on fund managers' investment decisions. Fund managers are prohibited from investing in ventures in which they have prior interests, such as ventures that they had previously invested in or currently advise through board memberships. This restriction prevents fund managers from rolling over investments across funds ("cross-fund" investment). Additionally, they are required to devote nearly all of their business time to managing a single fund, preventing them from simultaneously managing multiple funds with different sets of Limited Partners.[7] Some LPAs will specify that the fund managers be prohibited from raising the next fund until at least, for example, 75% of the fund's capital has been invested. Finally, the LPAs often specify an upper limit equal to 10 to 20% of the fund's value that can be invested into a single venture.

Partners generate a return on their investment when they sell, i.e. "exit" from, their equity investments (Kaplan and Stromberg, 2003; Gompers and Lerner, 2001; Berlin, 2014; Lerner and Nanda, 2020; Lerner, 2012). Fund managers are expected to distribute any income in proportion to the respective capital commitment by each partner but on aggregate, the savers usually receive 80% of the net profit. Fund managers are usually paid two percent of the total value of the fund annually as "management" fees to compensate for their time and investment expertise.

[6]The modern venture capital firm's organization can be traced to Draper, Gaither, and Anderson (DGA), a venture capital firm founded in 1959 in the Silicon Valley (Berlin, 2014). The DGA was organized as a Limited Partnership modeled after other investment organizations of the time such as the Rockefeller Brothers Fund and the J.H. Whitney Fund. The original DGA funds were closed-ended and had a five-year term. Since then, venture capital firms have continued to organize themselves as Limited Partnerships with term limits, although sometime in the 1970s, the dominant contract structure switched to a term of 10 years from five. Interviews of fund managers and savers reveal a "standard and inflexible" structure across the industry (Lerner and Nanda, 2020). This term structure is outlined in the model Limited Partnership Agreement (LPA) provided by the Institutional Limited Partners Association, with a sample term sheet reproduced in Appendix C. Similar partnership structures are found in other asset classes, such as buyout private equity, infrastructure, real estate, and natural resources (Brown et al., 2021).

[7]There is a technical exception to this requirement. Some individual Limited Partners may have statutory requirements that prevent them from receiving certain types of incomes, or that prevent them from holding a share of a venture capital fund that exceeds a threshold. These Limited Partners sign side contracts with the General Partners, but the LPAs stipulate that the investments from these funds happen concurrently and in proportion to the Limited Partner's share in the fund. These funds are often called Parallel Funds or Side-by-Side funds. These side contracts may be used to change the nature of fees charged to some Limited Partners (Begenau and Siriwardane, 2024).

**The Impact of Term Limits.** The prohibition on cross-fund investment affects managers' investment into long-term innovation. Investments whose value has not been realized before the fund's deadline cannot be simply rolled over to later funds with a different set of savers. In rare cases, fund managers are able to raise continuation funds with existing savers that allow them to roll-over specific investments. While continuation funds exist, they are rare in venture capital (Kastiel and Nili, 2024).

Managers could also exit from selling their equity shares in ventures to other managers ("secondaries"). However, as I describe in Section 6, secondaries are rare. Most exits happen when the venture is acquired by another company, goes out of business, or goes public.[8]

Another force pushes fund managers to invest in short-term innovation. As has been documented for other forms of private equity funds, managers whose funds perform well are more likely to raise more funds (Chevalier and Ellison, 1997; Barrot, 2017). This flow-performance relationship pushes fund managers to invest in ventures whose viability will be proven early so that they can raise follow-on funds faster. The LPAs, by preventing fund managers from investing out of more than one fund at a time, enhance this flow-performance relationship since managers must exhaust a prior fund before raising capital for the next.

# 3. Data

In this section, I outline the construction of the novel dataset used to evaluate the patience of venture capital. I begin by describing the data on innovation funding from venture capital firms, federal programs and agencies, and public firms. This is followed by an overview of the technologies covered in the dataset. I then briefly explain the procedure used to assign funding to these technologies. Detailed procedures for cleaning and merging the datasets can be found in. Detailed procedures for cleaning and merging the datasets are provided in Appendix A.

[8]Limited Partners may exit from the fund itself by selling their share of the fund to other Limited Partners. This is generally prohibited without the explicit permission of the General Partners; even then, the secondary market for venture capital fund shares is limited, though growing in the buyout industry (Nadauld et al., 2019).

## 3.1. Data on Innovation Funding

**Venture Capital Funding.** Data on venture capital comes from PitchBook, a database that has information on private capital transactions gathered through firm and fund contacts, news stories, and regulatory filings. PitchBook has been used extensively in academic articles on venture capital and also by the National Venture Capital Association and the U.S. National Science Board (Lerner et al., 2024). The dataset includes information on each funding round of the ventures which is each instance when a venture raises money from investors. Most funding rounds occur in syndicates with the involvement of several investors and I observe the names of investors, the date of the funding round, and the total amount of investment raised by the venture. I do not observe the contribution of each investor separately and follow the rest of the literature in distributing the total funding round amount equally amongst all investors. I define a "deal" as a combination of the funding round's identification number (ID), the investor's ID, and the ID of the investor's fund and restrict my sample to deals made between 1980 and 2022, inclusive, by investors whose primary type is listed by PitchBook as Venture Capital. I convert the total amount of each deal to 2022 constant dollars using data on CPI from the Federal Reserve Bank of St. Louis.

The final sample of venture capital data, as summarized in Table A10, includes data on 55,709 ventures, hundred and seven thousand funding rounds, and almost 2.2 trillion dollars of funding.

**Funding from Public Firms.** Data on research and development expenditures by public firms is taken from Compustat and adjusted to constant 2022 dollars as elsewhere. Public firms are the largest contributor to research and development expenditures; their total investment over the past four decades amounts to more than 19 trillion dollars, as summarized in Table A10. Part of the reason for the much larger investment from public companies compared to others is that the Compustat data includes each company's global expenditures on research and development.

**Funding from Federal Agencies and Programs.** Data for federal agencies and programs comes from agency and program-specific websites; details on the sources and cleaning decisions are presented in Appendix A.5. The objective is to collect information on the grants that were awarded for the purpose of research and development by these agencies between 1980 and 2022, including details on the type of innovation that was financed and

the funding amount.[9]

I collect information on two federal programs and four government agencies. The programs are the the Small Business Innovation Research program (SBIR) and the Small Business Technology Transfer (STTR). The agencies are the National Science Foundation (NSF), the National Institutes of Health (NIH, which are housed under the Department of Health and Human Services), the Department of Defense (DOD), and the Department of Energy (DOE). Appendix A.5 presents annual totals for the number of awards granted by each program and agency as well as the corresponding dollar amount in current dollars.

The final sample of funding from federal agencies is summarized in Table A10; the NIH is the largest contributor with an investment of almost 1 trillion dollars (in constant 2022 dollars) since 1980.

## 3.2. Defining Technologies

I use the Combined Patent Classification (CPC) system to construct the universe of technologies. The U.S. Patents and Trademarks Office (USPTO) uses the CPC system to categorize patents based on their technological content. The Guide to the classification system states that it has "the important purposes of serving as... a basis for investigating the state of the art in given fields of technology (WIPO, 2022)."

The third hierarchical level, or the 4-digit subclass, of the CPC forms my set of technologies. The correct choice of technological categorization trades off between two considerations: first, the divisions must not be too fine otherwise the text-matching procedure that I use to map funding to technologies will lead to several ambiguous matches. Second, the divisions must not be too aggregated or the comparison of technologies becomes meaningless. The 4-digit subclass balances these two considerations. Finally, choosing the subclass ensures that we can compare the technologies over time. In 2013, the USPTO moved from its old system of classification–the International Patent Classification (IPC) system–to the current CPC. However, the IPC and CPC systems correspond exactly at the 4-digit level. Using a finer level of division would necessitate mapping the older patents to the new system, which is complicated by the fact that several smaller categories were either grouped together or split apart.

[9] In practice, research and development correspond to different parts of the process of innovation (Arrow, 1993). Furthermore, the funding received by a venture or a research unit may be used for purposes other than innovation, for example, to pay for rent or travel of researchers. There is no clear way to isolate the funding that is spent on innovation so I do not make any further distinction.

There are 643 CPC subclasses based on this data. I drop CPC classes with fewer than 50 patents in my dataset; this leads to a final set of 584 CPC classes.

### 3.3. Assigning Funding to Technologies

I use natural language processing to allocate funding to technologies. I compile descriptions of technologies, ventures, public firms, and federal grants, and assign each venture, public firm, and grant to the technology with which its description is most similar. I now briefly describe this procedure.

The description of each CPC subclass combines the title and definition statements provided by the CPC. I obtain data on the CPC subclasses and their associated titles and definitions from the joint website of the European Patent Office and the United States Patent and Trademark Office. Each technology description has between one hundred to several hundred words. Appendix A.3 provides details on the construction of this dataset.

I first split the descriptions of CPC subclasses into words, or "tokens", using the Term Frequency-Inverse Document Frequency (TF-IDF) weights based on the vocabulary used in these descriptions. The TF-IDF procedure is used widely to tokenize texts (Kelly et al., 2021) and I implement it as follows. I standardize each technology description by removing punctuation and reducing ("lemmatizing") each word to its root form using large language models. The set of lemmatized words in the technologies forms my baseline vocabulary. For each word in this vocabulary, I calculate its relative importance as the inverse of the frequency with which it appears across descriptions (IDF). Words that appear frequently have smaller weights since they are less likely to be helpful in identifying technologies separately. I cast each technology description as a vector where each element of the vector is a term in the description, and its value is equal to the product of the term's frequency in that document and its IDF weight.

I also tokenize the descriptions of ventures, public firms, and federal awards using the IDF weights calculated from the technology descriptions. The descriptions of ventures comes from PitchBook, whose analysts use a standard template to describe each venture and highlight its unique technology or product (Lerner et al., 2024). Descriptions of Compustat firms are reported by Compustat under the *busdesc* variable. For awards from federal agencies, I use the titles and abstracts for the full set of awards granted since 1980.

Armed with a vectorized representation of each firm and technology description, I calculate the cosine similarity between each pair of vectors. I assign each venture, firm,

and abstract to the technology with which its most similar. In keeping with best practices, such as (Kelly et al., 2021), I eliminate matches where the cosine similarity is less than 5%. To assign venture capital funding to technologies, I aggregate the funding (in constant 2022 dollars) received by each firm from a venture capital investor and allocate it to the technological assignment of the firm.

Table A1 presents a validation check of this methodology. For each major industry group classification in PitchBook, I report the top two technology assignments by frequency. The Information Technology sector is mapped to several specific technologies such as: G06Q, which covers technologies related to information and communication technology, H04B, titled "transmission", which covers transmission of information-carrying signals, and G06V, which covers image and video generation. The table reflects the advantage of using CPC subclasses over the sector classification offered by PitchBook. Firms that are tagged to a broad technology class such as Information Technology are differentiated further based on the specific use-case. For example, the firm Kubos is described as a "space flight software platform." Rather than being clubbed with other firms building software, it is allocated to the technology class corresponding to Cosmonautics, B64G. The technology class A62B is linked to firms building medical devices specifically for respiration uses.

My method to use text-based similarity to allocate funding to technologies builds upon state-of-the-art techniques that have been recently used for similar purposes. For example, Autor et al. (2024) allocate patents to industries and occupations based on the similarity of the patent texts and industry descriptions. Similarly Bloom et al. (2021) detect the diffusion of new technologies across firms by measuring similarities between wikipedia pages describing technologies and firms' earnings calls. As with any statistical matching technique, this procedure introduces errors; some firm-technology links are false positives and some are false negatives, also called the precision-recall trade-off in machine learning. Appendix A.6 presents an overview of the alternative machine learning techniques I deployed and why the chosen procedure is the best in my setting.

# 4. Measuring Long-Term Innovation

To construct technology-level measures of gestation lag, I propose a statistical model of innovation that is relevant to the problem faced by fund managers when financing ventures.

I assume that each venture is an idea to improve upon a feature of an existing technology. For example, entrepreneurs working on batteries may have ideas related to a battery's flammability, the metals it uses, its weight, durability and so on. An idea needs investment upfront and only releases a public signal of its value after some time, which I assume is ex ante unknown and stochastic. I assume further that ideas related to a technology take a similar amount of time to be realized; I model this by assuming that the time between investment and a public signal of value for an idea is drawn from a technology-specific random variable. There are a fixed number of technologies in the economy denoted by the set $\mathcal{N}$, and the time to a public realization for an idea related to technology $j \in \mathcal{N}$ is drawn from an exponential distribution with a technology-specific rate parameter $\lambda_j$. The task is to estimate the expected time until public realization for an idea drawn from a technology $j$, $\tau_j \equiv 1/\lambda_j$. If I observe several realizations of this value for each technology, then I can estimate $\tau_j$ by calculating the sample average of these draws.

I take this model to the data by assuming that each patent in a CPC subclass corresponds to some feature of the technology and the time it takes to innovate upon this feature can be inferred from information on the patents that cite it. Table A11 summarizes the sample of patents that I use for this exercise. I restrict my sample to utility patents that were issued between 1980 and 2022 (inclusive), and for which at least one of the assignee is located in the U.S.[10] The latter restriction is in part because foreign assignees may have non-economic incentives that affect their choice of technology. For example, Akcigit et al. (2024) show that Chinese firms strategically invest in technologies in the U.S. that are more "basic" arguably to reach the frontier of these technologies faster. The other reason to restrict patents to those filed by U.S.-based assignees is that I focus on funding from sources that are most likely to target firms headquartered in the U.S. These include the U.S. federal government, venture capital firms headquartered in the U.S., and public firms in the U.S. I use PatentsView's assignee and location variables to make this determination. Filtering on the location of assignee implicitly drops patents that are not linked to any assignee, and the where the assignee is not linked to a location. The final sample of patents published in the U.S. between 1980 and 2022 and that have at least one U.S.-based assignee has 3,336,866 patents and is described in Table A11.

I link each patent in the final sample all patents that cite it; 23% of the patents not

[10] "Utility" patents are also called "patents for invention" and cover 90% of all granted patents (Lerner and Seru, 2022). I drop all non-utility patents and those that were tagged as "Withdrawn." I also drop patents whose IDs start with D, R, and P, which refer to Design, Re-issue, and Plant patents.

linked to any follow-on citation. For each citing patent, I observe and record its date of filing.[11] Appendix A.2 reports the definition and limits of what information is considered "prior art" and therefore needs to be cited by U.S. patents based on U.S. patent law (Title 35 of the United States Code.) The law states that a patent must cite *all* related patents that were filed before it and does not impose a statute of limitation such that patents filed before a certain date can be ignored.

I approximate the time to innovate upon a patent by the number of years between its publication date and the year when it receives half of all its future citations ("half-life"). To calculate a measure of gestation lag for the subclass, I calculate the average half-life for all patents assigned to that subclass. Section 5.4 presents a discussion of the robustness of this measure to various standard concerns regarding patents such as truncation of citations.[12]

Table A12 presents examples of CPC subclasses, their titles, the corresponding number of patents in those subclasses, and the estimated gestation lag for the top two (by count of patents) CPC subclasses in each quartile. The purple bars in Figure A4a report the distribution of gestation lags across these CPC subclasses; the gestation lag ranges from 2 to 13 years, with a median value of 7.2 years. The distribution of gestation lags across technologies in the economy is visually approximately normally distributed, with half of the technologies having gestation lags of greater than 7.2 years.

# 5. Evaluating the Patience of Venture Capital

In this section, I use the novel dataset described above to study whether venture capital is allocated to long-term innovation; I first compare its allocation to that by public firms, since public firms are a large, benchmark source of innovation funding in the private sector. I then infer the availability of commercial value and scientific opportunities within these technologies and compare their share of venture capital funding.

[11] The American Inventor Protection Act of 1999 mandated publication of "nearly" all patent applications filed on or after November 29, 2000 (Marx and Scharfmann, 2024).

[12] A brief comment on why it is appropriate to use different dates for the primary and the citing patents. When an entity files a patent application, the content of the application remain private until at least eighteen months have passed or the patent is granted. At the time it is granted, i.e. issued by the patent office, its contents are made public and the time between patent filing and granted dates can vary. Underlying the construction of the half-life measure is the assumption that follow-on innovation begins when the primary patent's technological contents become public information. Each patent is expected to cite every prior art it builds upon, without a cap on the age of the patents it must cite if they are relevant.

## 5.1. Venture Capital is Allocated to Short-Term Innovation

Figure 1a reveals that the distribution of venture capital resembles that by public firms. Public firms are a large source of funding for innovation; Table A10 reports that they spent 20 trillion USD in constant 2022 dollars on research and investment between 1980 and 2022, almost ten times the investment by the venture capital industry. More importantly, public firms are an important comparison group for venture capital since the design of the venture capital industry arguably faces none of the short-term pressures that public firms do such as quarterly earnings calls where they must meet earnings targets driven by career concerns. This figure establishes that despite the benefits of the venture capital industry, it seems unable to fund innovation that is longer term than public firms. In Figure A3, I restrict the sample to public firms that were not initially funded by venture capital firms using data from Ritter (2015). The distribution of these non-venture capital-backed public firms is identical to that of the rest of the sample.

Figure 1b compares the distribution of venture capital funding to the distribution of gestation lags in the economy. In purple, I plot the cumulative fraction of all the technologies in the U.S. economy with gestation lags measured on the x-axis. The cumulative fractions corresponding to venture capital are everywhere above the fractions of technologies, showing that venture capital is allocated to technologies with shorter lags than the technologies themselves. The median gestation lag, corresponding to fifty percent of the venture capital funding, is almost 2 years lower than the median gestation lag amongst technologies.

To show that these patterns are not simply driven by venture capital firms preferring one or two technologies with short gestation lags, in Figure A4a, I report histograms of venture capital and technologies. I bin both funding amounts (in 2022 constant dollars) and the technologies such that the bin at 3, for example, shows the fraction of technologies that have a gestation lag in the interval $(2, 3]$ and the fraction of total venture capital funding that is assigned to them. This figure shows that the concentration of venture capital into technologies with short gestation lags is not driven by just one or two technologies.[13]

[13]The Kullback-Leibler divergence between the allocation of venture capital and the distribution of gestation lags (based on half lives) in the economy is 0.53, which suggests the distributions are different from each other. In summary, the distribution of venture capital is shifted towards technologies with shorter gestation lags than the distribution of the universe of technologies in the U.S. economy.

## 5.2. Comparison with the Distribution of Commercial Value

Venture capital may be allocated to short-term technologies because there is a dearth of commercial value in the other technologies. In this section, I evaluate this possibility by computing the distribution of commercial value across technologies both directly–using patent-level measures of commercial value–and indirectly–by evaluating the allocation of two federal programs that are tasked with financing commercialization of technologies.

### 5.2.1. Distribution of Patent Value

I use data from (Kogan et al., 2017) who estimate commercial values of patents by exploiting high-frequency variation in stock prices around the announcement of patent grants. I calculate the sum of the commercial value in each CPC subclass and adjust it by the total number of patents. Figure 1c reports the cumulative distribution of the normalized commercial value across CPC subclasses. The allocation of venture capital is shifted substantially to the left of commercial value.[14]

Table 1 reports a univariate regression of inflation-adjusted venture capital funding, measured in billions, on the per-patent commercial value, also adjusted for inflation, gestation lag, and the interaction between the two variables. In the univariate regressions from columns 1 and 2, we learn that venture capital funding is more likely to be allocated to CPC subclasses with larger per-patent commercial value and less likely to be allocated to CPC subclasses with longer gestation lags. For every additional million dollar of normalized commercial value, the venture capital industry allocates 47 million dollars to it. Furthermore, every additional year of gestation lag reduces venture capital funding to the technology by 2 billion dollars. In column 3, I include both these controls as well as the interaction between them. The coefficient on the interaction term is significant; technologies with each additional year of gestation lag receive 50 million dollars less despite their commercial value. This table suggests that venture capital firms under-invest in

[14]To understand why it is reasonable to adjust the sum of commercial value in a subclass for the number of patents assigned to it, Figure A1a shows the distribution of the total sum of commercial value across the CPC subclasses, and Figure A1b reports the histogram of the distribution of the per-patent commercial value. While the total commercial value is concentrated in technologies that receive more venture capital, the per-patent commercial value is distributed exactly like the full distribution of technologies. One reason that the non-normalized distribution of commercial value resembles that of venture capital is that it is mechanically larger for technologies that have more patents. Normalizing by total count of patents in each technology allows a more accurate measure of the expected commercial value from undertaking innovation upon a technology. Other papers such as Kline et al. (2019) have also used the Kogan et al. (2017) values to infer commercial value of innovation.

technologies with long gestation lags, even if they have more commercial value.

**Table 1:** Link between commercial value in technologies and venture capital allocation

| | 1 | 2 | 3 |
|---|---|---|---|
| | Venture Capital Funding | | |
| Commercial Value (Per Patent) | 46.53*** | | 412.87*** |
| | (17.43) | | (60.44) |
| Gestation Lag | | -2282.52*** | 102.38 |
| | | (285.26) | (463.30) |
| Commercial Value X Gestation Lag | | | -49.56*** |
| | | | (7.82) |
| Intercept | 1878.69** | 19884.50*** | 254.57 |
| | (919.01) | (2056.63) | (3489.88) |
| R-squared | 0.01 | 0.10 | 0.17 |
| R-squared Adj. | 0.01 | 0.10 | 0.17 |
| Observations | 555 | 555 | 555 |

*Note:* This table shows the results from ordinary least squares regressions of venture capital funding allocated to technologies between 1980 and 2022, in millions of constant 2022 dollars, on three sets of explanatory variables. In the first column, technology-level venture capital funding is regressed on technology-level normalized commercial value. Commercial value of patents is also measured in millions of constant 2022 dollars. In the second column, technology-level venture capital funding is regressed on technology-level measures of gestation lag. The third column includes both these explanatory variables as well as their interaction. Robust standard errors in parentheses. ***, **, * indicate significance at 1%, 5%, and 10%, respectively. Details about the construction of these variables are presented in Section 3 and Section 4.

#### 5.2.2. Allocation of Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) Programs

An alternative way to infer the distribution of commercial value across technologies is to study the investment decisions of other innovation funding sources that share venture capital's objective to identify commercializable innovation. Two federal programs fit this bill: the Small Business Innovation Research Program (SBIR) and the Small Business Technology Transfer program (STTR). I provide an overview of the programs' regulatory mandates to show how their allocation may reveal the distribution of commercial value across technologies.

The SBIR is the largest single provider of grants for high-tech entrepreneurs in the U.S. (Howell, 2017). It was established under the Small Business Innovation Development Act

of 1982 with the goal to "strengthen the role of the small, innovative firms in federally funded research and development" and "stimulate technological innovation." The official website of the SBIR program (sbir.gov) calls it "America's Seed Fund," drawing a direct analogy to venture funding. Eleven federal agencies provide capital to the SBIR program: Department of Agriculture, Department of Commerce, Department of Defense, Department of Education, Department of Energy, Department of Health and Human Services, Department of Homeland Security, Department of Transportation, Environmental Protection Agency, National Aeronautics and Space Administration, the National Science Foundation, and the Small Business Administration. The projects that receive SBIR/STTR grants are selected by the federal agencies that provide funding to these programs. Each agency administers its own SBIR/STTR program and designates topics for research and development that it is seeking to fund. For a small business to be eligible for the SBIR program, it must have fewer than 500 employees, be organized as a for-profit organization, and be majority-owned by U.S. permanent residents or citizens. The program's objective, alongside its target set of small, young firms, makes its allocation directly comparable to that by venture capital firms.

The STTR program was established under the Small Business Technology Transfer Act of 1992. STTR is focused on opportunities where small businesses collaborate with U.S. research institutions. Five federal agencies provide funding to the STTR program: Department of Defense, Department of Energy, Department of Health and Human Services, and the National Science Foundation.

Despite the direct parallels between the SBIR program and venture capital, there are two major differences. First, the SBIR is tasked with funding commercialization of technology that directly meets the federal government's research and development needs, thus potentially limiting the gamut of innovations it can fund. Furthermore, its grant amounts are much more limited than the investments made by venture capital. SBIR program provides capital in two stages: Phase I and Phase II, where Phase I grants provide USD 150,000 and Phase II grants provide USD 1 million. The latter aims to fund later-stage demonstrations.

I combine the analysis of the SBIR and STTR programs because they are both intended towards small, innovative firms and are generally considered together by legislators. The SBIR/STTR programs have been evaluated previously and found to have a large effect on the subsequent innovation activities of firms that receive these grants (Lerner, 2000). Howell (2017) argues that despite the small amounts, the SBIR grants are crucial for inno-

vation because they allow constrained firms to develop prototypes, which later attracts venture capital funding. Since 1980, the programs have colletively distributed 90 billion dollars (in constant 2022 dollars) to awards.

Figure 3a compares the distribution of venture capital to the distribution by the SBIR/STTR programs. The cumulative distribution of SBIR/STTR lies below that of venture capital, suggesting that the SBIR/STTR programs are allocated to technologies with longer gestation lags compared to venture capital.

## 5.3. Comparison with the Distribution of Scientific Opportunities

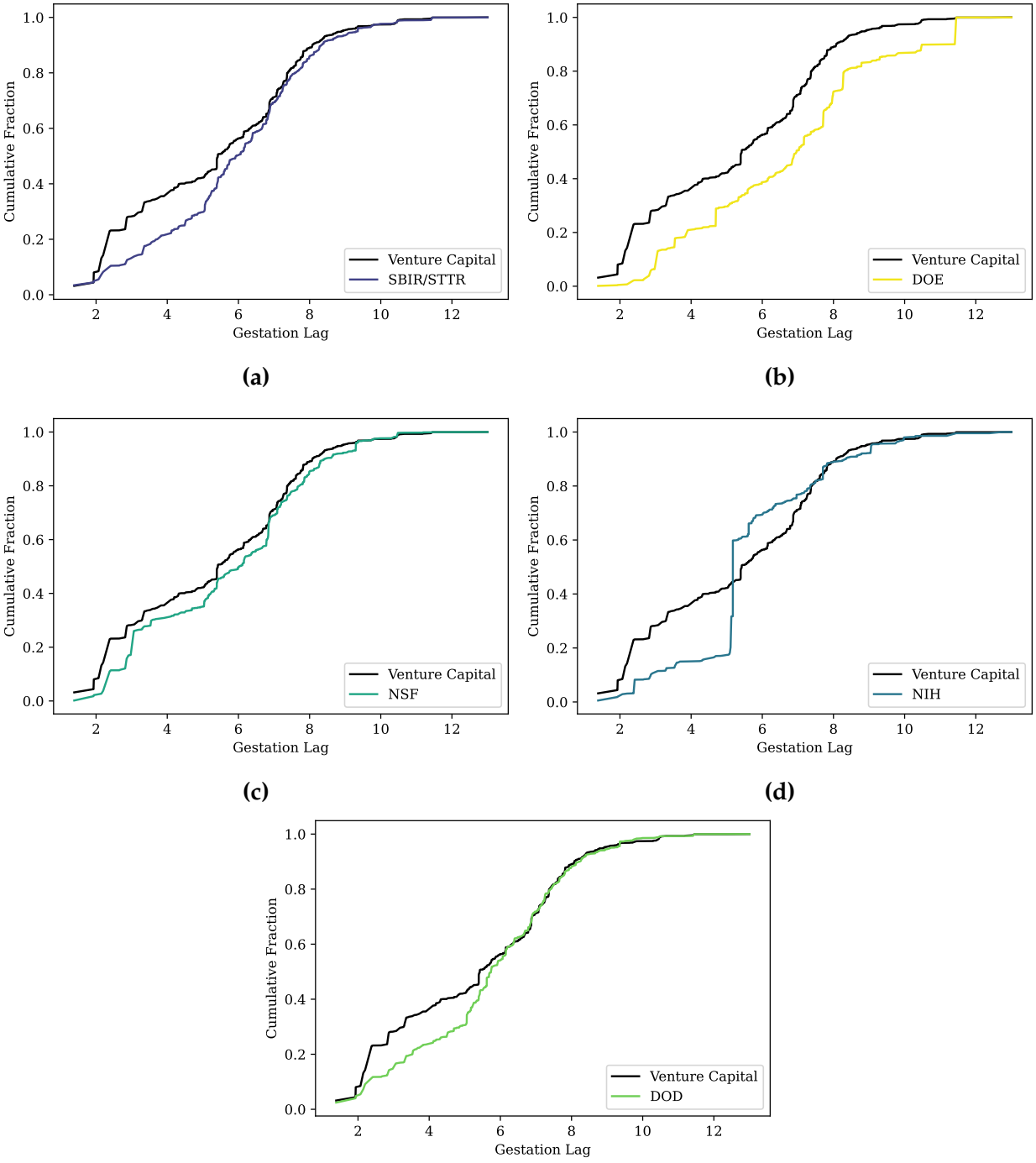
I evaluate whether there is less scientific opportunity in technologies that receive less venture capital funding by studying the allocations of funding by federal agencies responsible for most of the innovation in the U.S. economy.

I use the Government Patent Register (Gross and Sampat, 2024) to identify the federal agencies that are the largest contributors to innovation in the U.S. economy. The Government Patent Register is the most exhaustive dataset to date for the identification of patents that received federal government support. In keeping with the rest of the sample, I restrict the sample to patents granted between 1980 and 2022. Table A9 reports the fraction of patents attributable to each federal agency, where patents can be attributed to multiple agencies at once..[15]

The following agencies cumulatively account for ninety percent of the patents granted to the U.S. federal government in the past four decades: the Department of Defense (DOD), the Department for Health and Human Services that houses the National Institutes of Health (NIH), the Department of Energy (DOE), and the National Science Foundation (NSF). Panels (d), (e), and (f) of Figure 3 present the distributions of the DOD, the DOE, and the NSF, alongside the cumulative distribution of venture capital. Venture capital is allocated to more short-term technologies compared relative to each of these federal agencies.

[15]Note that the the Government Patent Register does not include SBIR/STTR programs in its list of federal funding sources since the funding for these programs is allocated directly via the responsible agencies. I have no easy way to tag patents that were funded by these programs.

**Figure 3:** Comparing Venture Capital to Federal Sources of Innovation Funding




*Notes:* Each panel reports the cumulative density function (CDF) of the distribution of venture capital against that of the other sources of funding as mentioned in the legends of the respective graph. To construct the CDF, I calculate the sum of all investments made by each source of innovation funding to each of the 584 technologies in my data after converting to constant 2022 dollars. Then, I arrange the technologies in the increasing order of their gestation lags, and calculate the fraction of funding allocated to each technology. The construction of the dataset underlying these figures is described in Section 3 and Section 4.

## 5.4. Robustness of the Gestation Lag Proxy

In this section, I show robustness of the half-life measure of gestation lags to various concerns regarding patent data.

First, I address the possibility that technologies with short gestation lags are those that have several recent patents with one or two cites. I limit the sample to patents that have at least 10 cites and where the patents were granted before 2018. For context, very few patents receive 10 cites; 10 citations correspond to the 70th percentile of patents' citation counts. Furthermore, by limiting the sample to patents granted before 2018, I address truncation, which is a perennial problem with using patents data. There is a delay between when an assignee files a patent and when it is granted by the patent office; Hall et al. (2001) and Dass et al. (2017) argue that patent-office lag can cause researchers to under-count patenting, as well as citations, for new firms and recently issued patents. While they focus on firm-level innovation, the truncation bias can show up also at the level of the technology if some technologies have experienced more innovation recently than others. Dass et al. (2017) argue that when dating patents according to their year of filing, excluding the final three to four years of patent data can address truncation bias. Figure A4b presents this analysis. The figure looks similar to Figure A4a, and in fact strengthens the conclusion that venture capital is distributed to technologies with short lags.

Second, I address the mechanical problem that older patents have had a longer time to accumulate citations, and will therefore have longer half lives. If venture capital is allocated to newer technologies that only have recent patents, rather than technologies with short gestation lags, our take-away will be misleading. I address this problem in several ways. First, I propose a statistical correction where I directly address for age differences in patents by weighing each patent's half-life by its age. Younger patents are given less weight because their full citation pattern is unlikely to be observed since they were published. This insight is similar to the corrections for length-biased sampling proposed in the bio-statistics literature (Heltshe et al., 2015). Figure A4c reports the age-weighted gestation lag. The take-away remains unchanged.

Then, I construct half-lives using the same window of 10 years for all patents. This limits the sample to patents granted between 1980 and 2013. Figure A4d presents the parallel figure using this alternative measure. The black bars continue to be shifted to the left relative to the purple bars, which means venture capital is allocated to technologies with shorter lags relative to the full distribution of technologies. Note, the x-axis on this

figure is changed to reflect that calculating half-lives within 10-year windows mechanically shortens the half lives. Instead of bins of 1 year, the bars reflect bins of 0.2 years, so the bin corresponding to 2.4 reflects all technologies with average half-lives between 2.3 and 2.4 years.

Third, I address the possibility of reverse causation, i.e., it is not that venture capital is being allocated to technologies with short gestation lags, rather, the allocation of venture capital is itself reducing gestation lags for technologies. I address this directly by exploring how correlated funding is to the shortening of gestation lag in CPC subclasses over time. I calculate the decline in half-lives of patents in each CPC subclass by regressing the annual mean half-life within a CPC class on the year. I then regress the estimated slopes on the total venture capital funding allocated to each subclass; Table A15 reports the results from this regression. The coefficient on venture capital funding is 0; there is no differential decline in the gestation lag of technologies that receive more venture capital funding.

While my paper is the first to attempt the measurement of gestation lags across all technologies in the economy, previous efforts have provided proxies for this measure for various industries. Appendix B.1 provides a short literature review of other measures in the literature and their similarity to mine. Table A14 presents the correlation of each extension of the half-life measure with the naive half-life average used in Figure 1. All these measures are highly correlated with the simplest one, so I continue to use it for the rest of this paper.

## 5.5. Correlates

As a society, is it important to invest in technologies with longer gestation lags? To understand what this novel measure of technological heterogeneity captures, I correlate it with five different measures from the literature that measure heterogeneity in innovation. The empirical proxy developed in this paper is correlated with most measures that attempt to capture the value of innovation–such as spillovers, or connection to basic science. However, the correlation is smaller than unity, suggesting that this measure is capturing a unique facet of innovation.

First, I use data from Marx and Scharfmann (2024) who construct patent-paper pairs to identify all patents that were closely built on academic papers. Patents that build directly on academic papers, or cite more academic papers, are considered to be more basic and have larger spillovers. For each technology, I calculate the fraction of patents in the CPC

subclass that are built directly on academic science and the number of days between the publication of scientific papers and related patents. Both these measures are meant to be increasing in the academic-ness or basicness of the technological field.

Second, I use patent-level measure of influence as constructed by Kelly et al. (2021). The KPST score proposed by this measure calculates a patent's influence on follow-on research based on the similarity between its text and the patents that follow. I use both the 1-year and the 5-year measures of influence and calculate the average importance of patents in each CPC subclass. Higher scores correspond to technologies which have more influential patents, which could happen if they provide fundamental innovation which is used by others.

Finally, I use the each patent's total citation count as another measure of quality. I demean each patent's total citations using the mean number of citations for patents granted in a given year.

Table A16 presents the correlation of each of these measures with the half-life measure. Column 1 presents correlation at the patent-level, and column 2 presents correlation at the level of the CPC subclass.

# 6. Evidence on the Effect of Term Limits

The empirical evidence in the previous section suggests that venture capital is disproportionately allocated to short-term innovation. In this section, I study the effect of the 10-year venture capital fund structure in driving this finding.

The key challenge in identifying the effect of term limits on the investment decisions of fund managers is that the ubiquity of the contract design renders no variation in term limits across funds. Instead, I compare the characteristics of ventures that receive investment from fund managers at different points in the fund's lifecycle. By comparing ventures that receive investment with 10 years left in the fund with, say, 9 years left in the fund, I approximate the decision what would be faced by a manager if she were facing a fund with a 10-year deadline or a 9-year deadline.

The null hypothesis is that term limits do not matter for investment decisions, which if true, would mean that the slope when regressing investment decisions on fund age is zero. This would happen if, for example, managers could costlessly roll-over their investments at the end of the fund's life or sell the investments to other investors without experiencing any losses. While cross-overs are usually prevented by LPAs, secondaries

are vanishingly rare, as I show using data on exits below. Note, further, that another feature of the contract helps us arrive at this null hypothesis; as discussed in Section 2, LPAs usually prohibit GPs from investing more than 10 to 15 percent of the fund's value in a single venture. Therefore, it is unlikely that the relationship between fund age and venture characteristics is driven purely by the amount of "dry powder", i.e., capital that is left in the fund at later stages of the fund's life.

I use two metrics to show that term limits matter for investment decisions. First is the age of the venture. Second is the time between the first investment and when the investor exits from the venture ("time to exit"). The first metric provides a venture-level measure of gestation lag.[16] The whole premise of this paper is that innovation takes time, and the time between when investment occurs and when its value is publicly realized is stochastic and ex ante unknown. It is reasonable, therefore, to think that younger ventures have, in expectation, a longer time until their value is realized relative to older ventures. Second, using time to exit as an outcome variable allows us to infer whether the deadline imposed by the LPA matters in practice. Here again the null hypothesis is that there is no relationship between fund age and the time that a fund manager stays invested in a venture.

**Fund-Level Data.** To study the relationship between fund age and investment decisions, I construct a dataset of first-time investments by venture capital funds into ventures using PitchBook. Since fund managers often make several, sequential investments into ventures, keeping all investments from fund managers into ventures will generate a mechanical relationship between fund age and venture characteristics. I restrict my sample to focus on funds that began between 1980 and 2013 (inclusive) and by investors whose primary type is "Venture Capital." These vintage boundaries ensure that I am able to observe the full life-cycle behavior of all funds in my analysis. In keeping with the rest of this paper, I restrict the sample to investors whose headquarters are in the U.S. Furthermore, I address the fact that most venture capital investments occur via syndicates by counting each investor-fund-funding round triplet as one deal (Kaplan and Stromberg, 2003). Hence, if there are two investors investing into a company at the same date, I count this as two deals. Additionally, if an investor invests into a company from two of

[16] We cannot use the technology-level measure of gestation lag in this analysis because there is little within-manager variation in the technologies they invest in. Other papers also find evidence of specialization amongst venture capital managers (Gompers et al., 2009).

his funds, I will count that as two separate deals.[17]

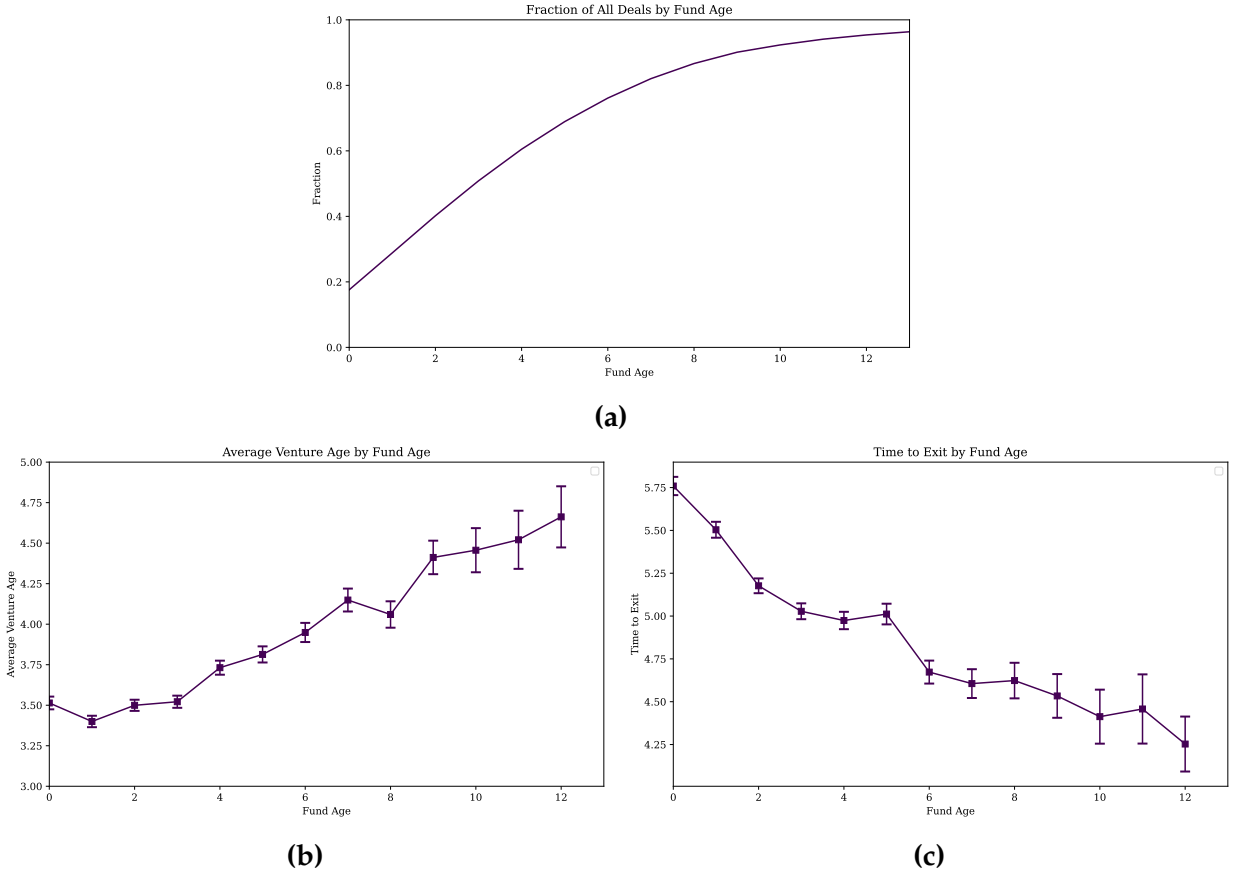
Panel A in Table A17 reports fund-level statistics. There are 3,508 funds in my dataset that meet these criteria. These funds make investment into 17 companies on average. The average age of the fund when making first-time investments into ventures is 3. There are 24,775 ventures in my sample and they are on average 3 years old when they receive their first venture capital investment. Panel C in Table A17 summarizes the data on first-time investments from fund managers into ventures. My sample has 62,205 first-time investments. 80% of these investments are made into companies when they are less than 5 years old; in turn, these investments are made by funds that are on average 4 years old. The average size of the deals is 8 million 2022 dollars; and about 7% of the deals are missing information on the size of the deal.

**Description of Investment Decisions over the Fund Lifecycle.** I first present an example of Kleiner Perkins, which is a major venture capital firm, to demonstrate that fund managers make first-time investments into ventures only in the first 3-5 years of a fund's lifecycle, as stipulated by the LPAs. Table A18 shows the number of deals made in each year by several of its consecutive funds. This pattern also holds in the aggregate; Figure 4a presents the fraction of all first-time investments made by venture capital funds between 1980 and 2022. About eighty percent of all deals are made within the stipulated 5-year period.

The bottom two panels of Figure 4 show the relationship between the two outcome variables of interest–venture age at first deal and time to exit–over the fund's lifecycle. In Panel (b), I present the average age of the venture, with bars representing the standard error. The x-axis reports the age of of the fund, which is calculated as the difference between the calendar year of the funding round and the initial close of the fund. Panel (c) reports the average and standard error of the number of years that the fund managers stay invested in a venture before exiting. Both variables exhibit a monotonic relationship with fund age. When investing earlier in the fund, managers choose ventures that are

[17] PitchBook is unable to report the individual contributions of each investor to a funding round because of the limitation of Form D, which is filed by ventures to the Securities and Exchange Commission (SEC) when issuing securities under exemptions granted by Regulation D. As reported in Table A10, the dataset on all venture capital funding rounds from 1980 and 2022 that I use for the aggregate facts contains information on 107,295 funding rounds. Of these, 30% are not linked to any fund. When restricting the sample to only include investments made by funds with vintages older than 2013, I necessarily eliminate funding rounds that are not linked to any funds. There are 47,845 funding rounds made between 1980 and 2013. Of these, there are 39,736 funding rounds where at least one of the investors was making a first-time investment into the venture.

**Figure 4:** Relationship between fund age and investment decisions




*Notes:* Panel (a) reports the cumulative fraction of all first-time investments made by funds into ventures against the age at which these investments were made. Panel (b) reports the average age of the venture by the age of the fund when the ventures receive their first-time investments from funds. Panel (c) reports the average time until exit for each first-time investment made by a fund into a venture. The bars in both Panels (b) and (c) report the standard errors of the means. Data underlying these figures is reported in Table A17 and discussed in Section 6.

younger. Furthermore, they stay invested in those ventures for at least 1.5 years longer than the ventures they invest in later.[18]

**Within-Fund Differences in Investment Decisions.** I quantify the relationship between the investment decisions and fund age, as displayed in univariate relationships in Figure 4, in a regression framework. I estimate the effect of each additional year in the fund's lifecycle on venture age and time to exit by using an ordinary least squares model. The resulting slope measures the sensitivity of the fund managers to term limits. For both outcomes, I implement the following regression:

$$Y_{i,t} = \beta_1 \text{FundAge} \in [2,3]_{i,t} + \beta_2 \text{FundAge} \in [4,5] + \beta_3 \text{FundAge} > 5 + \delta_i + \varepsilon_{i,t} \qquad (1)$$

Here $Y_{i,t}$ is the outcome measured for fund $i$ when investing at time $t$. $\text{FundAge} \in [2,3]_{i,t}$ is an indicator that the fund $i$'s age is 2 or 3 in year $t$, and similarly for $\text{FundAge} \in [4,5]$ and $\text{FundAge} > 5$. The omitted category is investments made in the first two years of the fund's life (since age starts at 0.) The regression includes fund fixed effects. Table 2 reports the estimates of $\beta_1$, $\beta_2$, and $\beta_3$ from these regressions

Column (1) of Table 2 reports estimates from this regression where the dependent variable is the age of the venture. Compared to the ventures that receive investment in the first two years of a fund's life, the ventures that receive investments in years 2 and 3 are six percent older. The few ventures that receive investment later, after year 5, are almost a whole year older. Similarly, Column (2) reports that fund managers spend 6 percent less time in ventures that receive their investment at years 2 and 3; and almost a whole year less in ventures that receive investment after age 5. Both these coefficients are highly statistically significant, allowing us to reject the null that the term limits do not matter for investment decisions.[19]

**Cross-Fund Investing and Secondaries are Rare.** Recall, term limits are unlikely to have an effect on fund managers' choice of when to invest and exit from ventures if they

[18] I present an alternative measure of a venture's gestation lag in Figure A6. At the time of a venture receiving a first-time investment from a fund, I count the number of funding rounds it had received previously. The figure shows a monotonic relationship between the number of previous deals received by a fund and fund age, which resembles the monotonic relationship between venture age and fund age. Since these two outcomes are so highly correlated, I only focus on venture age.

[19] An open question is whether these estimates can be aggregated up to fully account for the aggregate facts presented in the first part of the paper. I leave this exploration for future research.

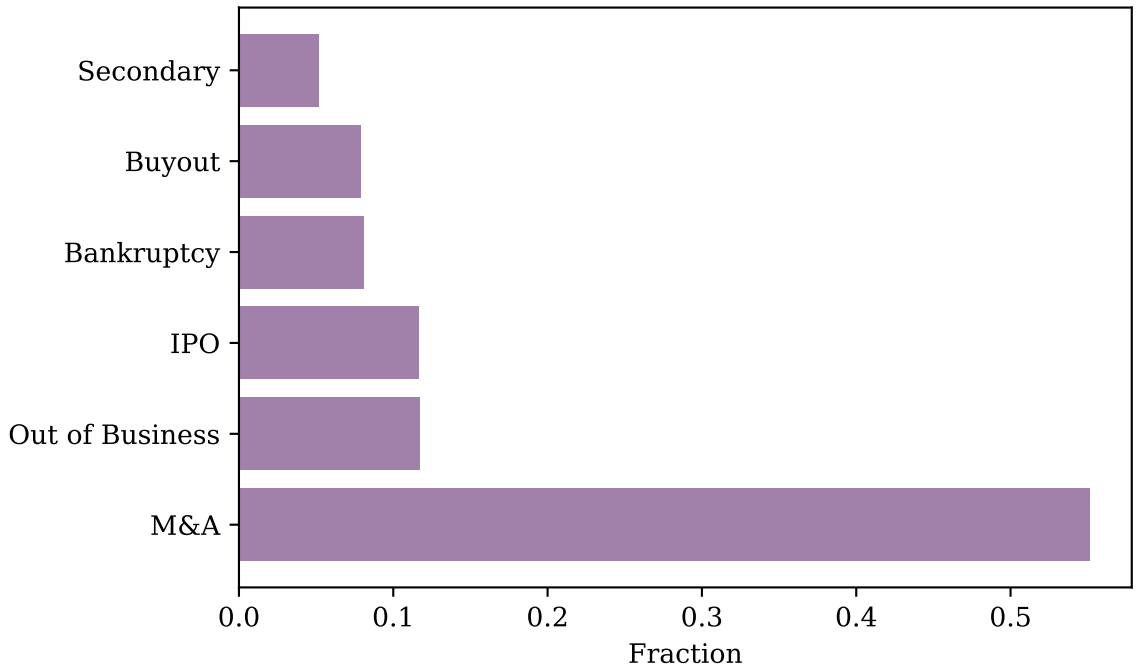
**Table 2:** Fund-Level Regression

| | (1) Venture Age | (2) Time to Exit |
|---|---|---|
| FundAge $\in$ [2,3] | 0.21*** | -0.35*** |
| | (0.04) | (0.06) |
| FundAge $\in$ [4,5] | 0.46*** | -0.55*** |
| | (0.06) | (0.07) |
| FundAge $>$ 5 | 0.81*** | -1.13*** |
| | (0.08) | (0.09) |
| Mean Dep. Var. FundAge$\in$ [0,1] | 3.45 | 5.63 |
| Fund FE | Yes | Yes |
| R-squared | 0.21 | 0.15 |
| R-squared Adj. | 0.15 | 0.06 |
| N | 61,856 | 43,586 |

*Notes:* Each column reports estimates from a separate regression. In column (1), I regress the age of ventures that receive a first-time investment from a fund on the fund's age. In column (2), the dependent variable is the time between investment and exit for all first-time investments made by funds. Fund age is measured from the initial closing. All standard errors are clustered at the level of the fund. Data underlying these regressions is discussed in Section 6. ***, **, * indicate significance at 1%, 5%, and 10%, respectively.

can costlessly roll-over their investments or sell them to other fund managers. Figure 5 shows the different ways in which fund managers exit from ventures in my data. The most infrequent way to exit from startups is a secondary sale to other investors; the share of secondary sales is less than 5 percent. The most frequent way to exit is through a merger and acquisition, which occurs in 55 percent of the cases. [20]

**Figure 5:** Exits in the Data




*Notes:* This table reports the fraction of each type of exit in the data for each first investment made by a fund into a company. Data underlying this series is described in Appendix A.4.

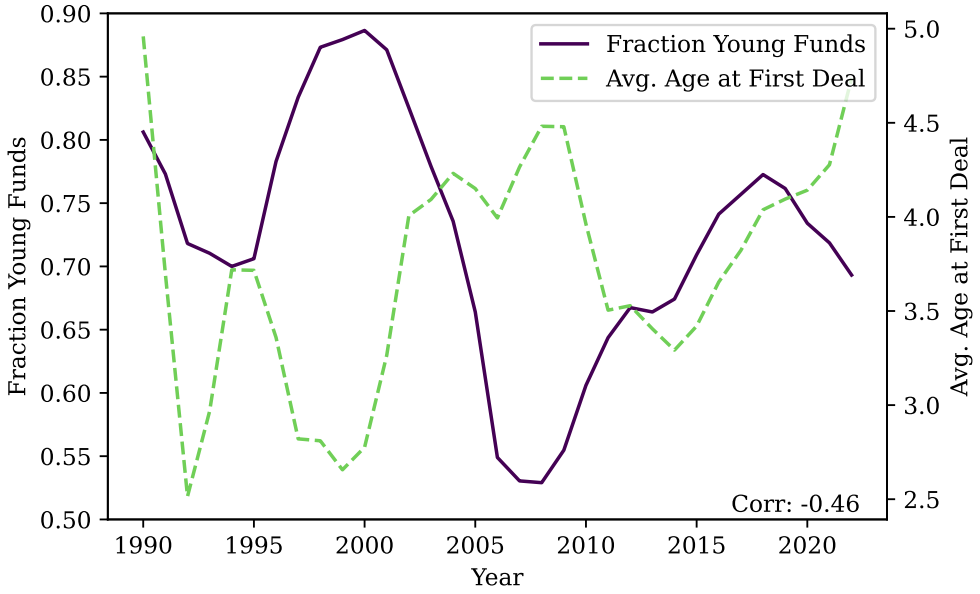
## 6.1. Instrumental Variable: Accounting for the Endogeneity of Availability of Young Funds

So far, I have shown that the remaining time in the fund affects managers' investment decisions. However, more funds may be raised in the same year that more ventures enter,

[20] Admati and Pfleiderer (1994) provide a theoretical rationale for why early investors into ventures do not fully exit from their investments until the venture has reached a "final" outcome like being acquired by another company, listing on the public stock exchange, or going out of business. They differentiate between early and later investors as "inside" and "outside" investors and argue that if inside investors were to fully exit from the investment, outside investors would infer that the venture was of low quality.

which would create a mechanical relationship between the age of the fund and the age of the ventures that the fund managers invest in. To address this endogeneity, I turn to aggregate data and study the relationship between the availability of young funds that are still in their investment period and investment into young ventures. I propose an instrument for the availability of young funds to isolate the savings that flow into venture capital for reasons that are arguably exogenous to availability of investment opportunities in ventures. Using this instrumental variable, I show that venture capital is invested into younger ventures when funds are younger.

**Figure 6:** Availability of Young Funds and Age of Ventures at Investment




*Notes:* To calculate the fraction of funds in each year that are in their investment period–"young funds"–I assume that each fund lasts for 10 years and use funds' vintages to mark their starting year. The average age of ventures is calculated when they receive their first investment from a fund. The data underlying these series is described in Section 6.

Figure 6 plots the fraction of funds that are young in a given year (in the solid purple line) and the average age of ventures when they receive their first investment from a fund (in the dashed green line.) I construct the annual availability of young funds by assuming that every fund is active for a period of 10 years and calculating the fraction of funds that are active but within their investment period, i.e., within 5 years since initial closing. To construct this series, I use all the funds that were raised between 1980 and 2022 (inclusive). There is a large variability in the availability of young funds across years: in

1999, almost 90% of all available funds were young. This collapsed to fewer than 60% in the aftermath of the 2000 financial crisis. There is also similarly large variability in the age or maturity of ventures when they receive a first investment from funds. When most of the funds were young, just before 2000, the average age of ventures receiving a first-time investment was about 2.5, about half of the age of ventures receiving capital in 2006. The correlation between these two series is -0.46.

The relationship between the availability of young funds and investment into young ventures is endogenous; if innovation opportunities arise in some technologies, then more entrepreneurs will start new ventures, and more savings will flow into the venture capital industry. This is certainly true for the rapid increase in funding in the late 1990s, when ventures using the internet were entering the market (Townsend, 2015) I therefore seek an instrument that can shift the availability of young funds but is unrelated to the availability of innovation opportunities.

I address the endogeneity captured in Figure 6 by instrumenting for the availability of young funds with exogenous increases in supply of capital to venture capital firms. To do this, I follow an insight in Brown et al. (2021) who argue that it is difficult for savers to time exposure to investment opportunities in venture capital because of the design of the venture capital funds as well as regulatory requirements on the Limited Partners that specify how exposed they can be to venture capital. Raising a new fund can take 2 to 3 years, and the standard contract structure gives Limited Partners no say in when the General Partners allocate capital to ventures (Janeway et al., 2021). The threat that Limited Partners correctly recognize the availability of new investment opportunities in a technology in a year, and are able to immediately find a General Partner who can allocate capital to ventures in that year, is structurally limited. Furthermore, several types of Limited Partners–such as pension funds and endowments–are statutorily required to limit their investment in a risky asset such as venture capital. Therefore, they can only expand their investments into venture capital when the total value of their portfolio increases for reasons unrelated to investment opportunities in venture capital.

Using this insight, I use the 5-year returns in the S&P 500 constructed from the inflation-adjusted Shiller Total Return P/E ratio as my instrument.[21] Figure A7a and Figure A7b report the first stage and the reduced form using this instrument, respectively. The instrument is valid if 5-year returns in the public equities market in year $t$ are independent of the contemporaneous investment opportunities in innovation. Table 3 reports the results

[21] The returns are constructed as of January 1 in each year.

from the two-stage least squares regression using 5-year returns as an instrument for the fraction of young funds in the economy. Column (1) in Table 3 reports results from the endogenous regression of the age at which ventures receive investment on the fraction of young funds. Columns (2) and (3) in Table 3 report the first stage and reduced form regressions respectively, and Column (4) reports the instrumental variable regression. I find that a 10 percentage point increase in the fraction of young funds reduces the average age of ventures receiving a first-time investment by 4 years. This relationship is highly significant.

**Table 3:** Availability of Young Funds and Investment in Young Ventures

| | (1) Venture Age | (2) Frac. Young Funds | (3) Venture Age | (4) Venture Age |
|---|---|---|---|---|
| Frac. Young Funds | -2.85*** | | | -4.44*** |
| | (0.94) | | | (1.61) |
| 5-Year Return | | 0.18*** | -0.78** | |
| | | (0.03) | (0.31) | |
| Intercept | 5.78*** | 0.69*** | 3.87*** | 6.93*** |
| | (0.64) | (0.02) | (0.09) | (1.17) |
| R-squared | 0.21 | 0.42 | 0.21 | 0.14 |
| R-squared Adj. | 0.18 | 0.40 | 0.18 | 0.12 |
| Observations | 33 | 33 | 33 | 33 |

*Notes:* Each column reports estimates from a separate regression. Column (1) reports estimates from an ordinary least squares regression of average age of the venture that receives a first-time investment from a venture capital fund on the fraction of funds that are young in a year. This is the main endogenous regression of interest. Column (2) reports the first stage, where I regress fraction of young funds in a year on the 5-year S&P return. Column (3) reports the reduced form from a regression of average age of ventures that receive funding in a year on the 5-year S&P return. The last column reports estimates from a two-stage least squares regression which uses 5-year returns as an instrument for the availability of young funds. Heteroskedasticity and autocorrelation robust standard errors are reported in parentheses with 8 lags. Significance of the coefficients is robust to choosing other levels of lags such as 5. ***, **, * indicate significance at 1%, 5%, and 10%, respectively.

## 6.2. Instrumental Variable: Market Sensitivity of Limited Partners

Another way to study the relationship between the availability of young funds and investment into young ventures is to compare, cross-sectionally, investors who raise different number of funds. To conduct this exercise, I focus on the period following the crash of the dot-com bubble (2000-2007) and study the relationship between investment in young ventures and the number of funds that venture capital firms raised in this period. Since

venture capital firms may be able to raise more funds if they have access to younger ventures, I propose a second instrument that exploits differences across venture capital firms in their exposure to Limited Partners with varying market sensitivities. By comparing the investment decisions in the same time period of venture capital firms that are able to raise different number of funds due to the institutional nature of their Limited Partners, I control for the possibility that availability of investment opportunities drives the relationship between availability of funds and investment in younger ventures.

I distinguish venture capital firms based on whether they received investments from endowments or pension funds before 1999. Endowments are considered to be a more patient source of capital than others, but especially relative to pension funds that face short-term pressures (Swensen, 2000; Ivashina and Lerner, 2019). I confirm that endowments are less sensitive to contemporaneous market conditions than public firms in Table 4. In the first column, I report estimates from a regression of the annual growth rate in the number of funds invested in by pension funds on the average quarterly returns in venture capital in that year. The second column reports estimates from the same regression but estimated on the sample of endowments. The market sensitivity of pension funds is almost twice that of endowments. To confirm that this difference is significant, I run a pooled regression on a sample with both types of Limited Partners; in the third column, the coefficient on the interaction term shows that endowments are significantly less reactive to contemporaneous returns than pension funds.[22]

Having established that endowment are less market sensitive than pension funds, I compare venture capital firms that had an investment relationships with endowments versus pension funds prior to 1999. The first column in Table 5 reports estimates from a regression of the average age of ventures on the number of funds raised by a venture capital firm between 2000 and 2007. The second column reports the first stage of this regression; venture capital firms that raised investment from endowments before 1999 raise almost 1.5 additional funds between 2000 and 2007 relative to firms that raised capital from pension funds. Column (3) reports the reduced form; venture capital firms that raised

[22]Data on the institutional type of Limited Partners comes from PitchBook. While PitchBook reports detailed institutional types, I aggregate them into three categories; Pension Funds, Endowments, and Other. Table A19 summarizes the count of funds associated with each category. Column 1 reports a coarse institutional type; column 2 reports the detailed institutional type reported by PitchBook, and column 3 reports the number of funds each institutional type invests in between 1980 and 2022. I do not use the third category, Other, in any of the analyses. Figure A8 provides an alternative way to visualize the market sensitivity of pension funds relative to endowments. The black solid line plots the average number of new funds (by vintage) that an LP invests in for each year; compared to the dashed grey line, which represents this average for endowments, pension funds both show a much larger run up before 2000 and a much larger crash.

**Table 4:** Market Sensitivity of Limited Partners

| | (1) Pension | (2) Endowment | (3) Combined |
|---|---|---|---|
| VC Return | 1.72*** | 0.94*** | 1.72*** |
| | (0.24) | (0.33) | (0.24) |
| Endowment | | | 3.02 |
| | | | (4.21) |
| VC Return X Endowment | | | -0.77* |
| | | | (0.41) |
| Intercept | -4.45 | -1.43 | -4.45 |
| | (3.50) | (2.07) | (3.50) |
| R-squared | 0.25 | 0.13 | 0.21 |
| R-squared Adj. | 0.22 | 0.10 | 0.16 |
| Observations | 28 | 28 | 56 |

*Notes:* Each column reports estimates from a separate regression. In the first column, I regress the annual growth in the number of funds that an average pension fund invests in on the average quarterly returns in venture capital for that year. The row corresponding to *VC Return* reports the estimated coefficient on the growth rate in venture capital returns. In the second column, I report the same coefficient for endowment funds. The last column combines the samples from the first two columns and regresses the growth in the average number of funds raised by each type of limited partner on venture capital returns as well as an interaction between the return and whether the limited partner is an endowment. Through these regressions, we learn that the beta for Pension funds is almost twice as large as the beta of endowments and this difference is statistically significant. Data on venture capital returns is obtained as part of the State Street Private Equity Index. The returns in the raw data are reported quarterly; I average the quarterly returns in a year to calculate the average growth rate in venture capital returns in that year. The data is available starting 1996; The sample is annual ranging from 1996 to 2024.

investments from endowments invest in ventures that are about half a year younger between 2000 and 2007. The last column reports estimates from a two-stage least squares regressions where the instrument is a binary indicator that equals 1 if the venture capital firm raised capital from endowments before 1999 and 0 otherwise.

# 7. Stickiness of Contracts

Why does the 10-year fund structure prevail in the venture capital industry? Answering this question theoretically is beyond the scope of this paper because of the following ob-

**Table 5:** Using Endowment LPs as IV

| | (1) Venture Age | (2) Num Funds | (3) Venture Age | (4) Venture Age |
|---|---|---|---|---|
| Num Funds | -0.08* | | | -0.28 |
| | (0.05) | | | (0.22) |
| 1=Endowment, 0=Pension | | 1.44*** | -0.41* | |
| | | (0.53) | (0.24) | |
| Intercept | 4.42*** | 3.07*** | 4.25*** | 5.12*** |
| | (0.27) | (0.20) | (0.17) | (0.77) |
| R-squared | 0.02 | 0.07 | 0.01 | -0.08 |
| R-squared Adj. | 0.01 | 0.06 | 0.01 | -0.09 |
| Observations | 134 | 134 | 134 | 134 |

*Notes:* Each column reports estimates from a separate regression. In Column (1), I regress the average age of ventures that receive investment from venture capital firms between 2000 and 2007 on the number of funds that these firms raise in this time. The age of the venture is calculated at the time of the first investment from the firm into the venture. In column (2), I regress the number of funds on the treatment indicator, where the indicator is 1 if the venture capital firm raised capital from an endowment in 1999 and 0 if it raised capital from a pension fund. If a venture capital firm raised capital from both types of institutions, I assign them to treatment, i.e., that they raised capital from endowments. Column (3) reports the reduced form regression of venture age on treatment status. Column (4) reports estimates from a two stage least squares regression where I instrument for the number of funds that a venture capital firm raised between 2000 and 2007 by the institutional type of limited partner it raised capital from in 1990. Robust standard errors are reported in parentheses.

servation: the venture capital industry is not unique in its widespread adoption of default contract terms. The adoption of boilerplate contract language has been documented extensively across industries and especially in finance (Gulati and Scott, 2011). In an evaluation of half a million contracts filed with the Securities and Exchange commission, Nyarko (2021) finds that a crucial clause, the dispute resolution clause, remains unspecified even between highly sophisticated parties. Gulati and Scott (2011) document that a clause had remained uninterrogated for over a century in contracts governing cross-country financial transactions. Both of these papers, and others, provide case studies where adopting the default provisions proved highly costly for relevant parties. Therefore, a theory that explains the adoption of defaults cannot rely on the specifics of any industry.

The legal literature has put forth several hypotheses for why default provisions may prevail even when potentially costly and I highlight one which seems most relevant in my setting. While canonical models of optimal contracts assume that contracts are a product of negotiation between two parties–a principal and an agent–these papers highlight the important role of a third agent: the law firm. Richman (2011); Kahan and Klausner (1997)

argue that large law firms have an interest in creating and using default templates due to economies of scale as well as the design of the legal system that relies upon precedence and certainty, leading to network externalities and path dependence. Richman (2011) argues that the writing of a contract rarely begins with a specific agent; rather, since a law firm has multiple clients and acquires information about an industry over time, using standardized templates "filter[s] out potentially damaging discretion." Nyarko (2021) and Kahan and Klausner (1997) find that a large portion of the variation in contract provisions across parties can be explained by the law firms they use.[23]

# 8. Conclusion

This paper investigates whether venture capital firms invest in long-term innovation. Using patent data, I propose an empirical proxy to measure the expected lag between investment and returns when innovating on a technology, which I define as the technology's gestation lag. I find that the allocation of funding to long-term innovation by venture capital firms mirrors the allocation by public firms. I also show that the concentration of venture capital firms to short-term innovation cannot be fully explained by those technologies having fewer scientific opportunities or less commercial value. To understand why there may be short-term pressures in the venture capital industry, I examine the incentives created by the ubiquitous 10-year fund structure, which encourages fund managers to invest in innovation whose value is realized early. Using fund-level data, I show that term limits significantly affect investment decisions.

This paper contributes to the literature in the economics of innovation that seeks to document and understand frictions in the private sector financing of innovation (Budish et al., 2015; Bloom et al., 2019). While prior studies have focused on the role of patents in both encouraging and limiting innovation, I shine light on voluntarily adopted contracts in the venture capital industry—a sector that plays a dominant role in financing innovation in the U.S. economy. Given the dual trends of the rising importance of the private sector in financing innovation and declining productivity growth, understanding

[23]Lastly, the organizational economics literature has also observed and remarked upon simplicity in real-world contracts that departs from the informationally sensitive contracts prescribed by mathematical models. Hart and Holmstrom (1989) note that "Real world schemes are simpler than theory would dictate and surprisingly uniform across a wide range of circumstances." Gottlieb and Moreira (2022) provide one theoretical defense of simple models. However, there is a subtle but important difference between the adoption of simple contracts and the prevalence of default terms. I believe little formal modeling has addressed either the role of the law firm or the efficiency of adopting contract defaults.

other institutional constraints that prevent the private sector from efficiently financing innovation is a promising area for future research (Bloom et al., 2020; Babina et al., 2023)

Another avenue for future research is measuring private and social rates of returns for investment in innovation. While a large body of literature has attempted to measure social returns to innovation, such as the early contribution by Griliches (1958), we have limited evidence on private rates of return and how they might change as private sector investment into innovation expands.

# References


Acemoglu, D. (2011). Diversity and Technological Progress. Working Paper 16984, National Bureau of Economic Research.

Acemoglu, D., P. Aghion, L. Bursztyn, and D. Hemous (2012). The Environment and Directed Technical Change. *American Economic Review 102*(1), 131–166.

Acemoglu, D., U. Akcigit, and M. A. Celik (2022). Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators. *American Economic Journal: Macroeconomics 14*(3), 199–249.

Admati, A. R. and P. Pfleiderer (1994). Robust Financial Contracting and the Role of Venture Capitalists. *The Journal of Finance 49*(2), 371–402.

Akcigit, U., S. T. Ates, J. Lerner, R. R. Townsend, and Y. Zhestkova (2024). Fencing Off Silicon Valley: Cross-Border Venture Capital and Technology Spillovers. *Journal of Monetary Economics 141*, 14–39.

Akcigit, U. and W. R. Kerr (2018). Growth Through Heterogeneous Innovations. *Journal of Political Economy 126*(4), 1374–1443.

Arora, A. and M. Dell (2024). LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models. Technical report.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. Princeton University Press.

Arrow, K. J. (1993). Innovation in Large and Small Firms. *The Journal of Entrepreneurial Finance 2*(2), 111–124.

Asker, J., J. Farre-Mensa, and A. Ljungqvist (2015). Corporate Investment and Stock Market Listing: A Puzzle? *The Review of Financial Studies 28*(2), 342–390.

Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu (2020). Foreign Competition and Domestic Innovation: Evidence from US Patents. *American Economic Review: Insights 2*(3), 357–374.

Autor, D., A. Salomons, and B. Seegmiller (2024). New Frontiers: The Origins and Content of New Work, 1940–2018. *The Quarterly Journal of Economics 139*(3), 1399–1465.

Azoulay, P., J. S. G. Zivin, and G. Manso (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics 42*(3), 527–554.

Babina, T., A. X. He, S. T. Howell, E. R. Perlman, and J. Staudt (2023). Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneur-

ship, and Publications. *The Quarterly Journal of Economics 138*(2), 895–954.

Banerjee, A. V. and E. Duflo (2014). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *The Review of Economic Studies 81*(2), 572–607.

Barrot, J.-N. (2017). Investor Horizon and the Life Cycle of Innovative Firms: Evidence from Venture Capital. *Management Science 63*(9), 2773–3145.

Bastianin, A., P. Castelnovo, M. Florio, and A. Giunta (2022). Big Science and Innovation: Gestation Lag from Procurement to Patents for CERN Suppliers. *The Journal of Technology Transfer 47*(2), 531–555.

Begenau, J. and E. N. Siriwardane (2024). Fee Variation in Private Equity. *The Journal of Finance 79*(2), 1199–1247.

Berlin, L. (2014). The First Venture Capital Firm in Silicon Valley: Draper, Gaither and Anderson. In *Making the American Century: Essays on the Political Culture of Twentieth Century America*. Oxford University Press.

Bernstein, S. (2015). Does Going Public Affect Innovation? *The Journal of Finance 70*(4), 1365–1403.

Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun (2021). The Diffusion of Disruptive Technologies. Working Paper 28999, National Bureau of Economic Research.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020). Are Ideas Getting Harder to Find? *American Economic Review 110*(4), 1104–1144.

Bloom, N., J. Van Reenen, and H. Williams (2019, August). A Toolkit of Policies to Promote Innovation. *Journal of Economic Perspectives 33*(3), 163–184.

Brown, G., R. Harris, W. Hu, T. Jenkinson, S. N. Kaplan, and D. T. Robinson (2021). Can investors time their exposure to private equity? *Journal of Financial Economics 139*(2), 561–577.

Bryan, K. and H. Williams (2021). Innovation: Market Failures and Public Policies. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), *Handbook of Industrial Organization, Volume 5*, pp. 281–388.

Bryan, K. A. and J. Lemus (2017). The Direction of Innovation. *Journal of Economic Theory 172*, 247–272.

Budish, E., B. N. Roin, and H. Williams (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review 105*(7), 2044–2085.

Buera, F. J., J. P. Kaboski, and Y. Shin (2015). Entrepreneurship and Financial Frictions: A

Macrodevelopment Perspective. *Annual Review of Economics 7*(1), 409–436.

Chevalier, J. and G. Ellison (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy 105*(6), 1167–1200.

Choi, S., M. Gulati, and R. Scott (2017). The Black Hole Problem in Commercial Boilerplate. *Duke Law Journal 67*(1), 1–77.

Dass, N., V. Nanda, and S. C. Xiao (2017). Truncation Bias Corrections in Patent Data: Implications for Recent Research on Innovation. *Journal of Corporate Finance 44*, 353–374.

Diamond, D. W. (1989). Reputation Acquisition in Debt Markets. *Journal of Political Economy 97*(4), 828–862.

Dyevre, A. (2023). Public R&D Spillovers and Productivity Growth. Technical report.

Fieldhouse, A. J. and K. Mertens (2023). The Returns to Government R&D: Evidence from U.S. Appropriations Shocks. Working Paper 2305, Federal Reserve Bank of Dallas.

Garfagnini, U. and B. Strulovici (2016). Social Experimentation with Interdependent and Expanding Technologies. *The Review of Economic Studies 83*(4), 1579–1613.

Gompers, P., A. Kovner, and J. Lerner (2009). Specialization and Success: Evidence from Venture Capital. *Journal of Economics & Management Strategy 18*(3), 817–844.

Gompers, P. and J. Lerner (2001). The Venture Capital Revolution. *Journal of Economic Perspectives 15*(2), 145–168.

Gompers, P. A. (1996). Grandstanding in the Venture Capital Industry. *Journal of Financial Economics 42*(1), 133–156.

Gottlieb, D. and H. Moreira (2022). Simple Contracts with Adverse Selection and Moral Hazard. *Theoretical Economics 17*(3), 1357–1401.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The Economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics 40*(1), 3–73.

Griliches, Z. (1958). Research Costs and Social Returns: Hybrid Corn and Related Innovations. *Journal of Political Economy 66*(5), 419–431.

Gross, D. P. and B. N. Sampat (2024). The Government Patent Register: A New Resource for Measuring U.S. Government-Funded Patenting. Technical Report 32136, National Bureau of Economic Research.

Gulati, G. M. and R. E. Scott (2011). The Three and a Half Minute Transaction: Boilerplate and the Limits of Contract Design. *Hofstra Law Review 40*(1), 1–9.

Hall, B. H., Z. Griliches, and J. A. Hausman (1986). Patents and R & D: Is There a Lag? *International Economic Review 27*(2), 265–283.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. Working Paper 8498, National Bureau of Economic Research.

Hart, O. and B. Holmstrom (1989). The Theory of Contracts. In T. F. Bewley (Ed.), *Advances in Economic Theory: Fifth World Congress*, pp. 71–156. Cambridge.

Heltshe, S. L., K. Kafadar, and P. C. Prorok (2015). Quantification of Length-Bias in Screening Trials with Covariate-Dependent Test Sensitivity. *Biometrical Journal. Biometrische Zeitschrift 57*(5), 777–796.

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

Hopenhayn, H. and F. Squintani (2021). On the Direction of Innovation. *Journal of Political Economy 129*(7), 1991–2022.

Howell, S. T. (2017). Financing Innovation: Evidence from R&D Grants. *American Economic Review 107*(4), 1136–1164.

Ivashina, V. and J. Lerner (2019). *Patient Capital: The Challenges and Promises of Long-Term Investing*. Princeton University Press.

Jaffe, A. B., M. S. Fogarty, and B. A. Banks (1998). Evidence from Patents and Patent Citations on the Impact of NASA and other Federal Labs on Commercial Innovation. *The Journal of Industrial Economics 46*(2), 183–205.

Janeway, W. H., R. Nanda, and M. Rhodes-Kropf (2021). Venture Capital Booms and Start-Up Financing. *Annual Review of Financial Economics 13*(1), 111–127.

Jones, B. F. and L. H. Summers (2021). A Calculation of the Social Returns to Innovation. In A. Goolsbee and B. F. Jones (Eds.), *Innovation and Public Policy*. University of Chicago Press.

Kahan, M. and M. Klausner (1997). Standardization and Innovation in Corporate Contracting (Or "The Economics of Boilerplate"). *Virginia Law Review 83*(4), 713–770.

Kaplan, S. and P. Stromberg (2003). Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts. *Review of Economic Studies 70*, 281–315.

Kastiel, K. and Y. Nili (2024). The Rise of Private Equity Continuation Funds. Working Paper 340, Stigler Center for the Study of the Economy and the State.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights 3*(3), 303–320.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who Profits from Patents? Rent-

Sharing at Innovative Firms. *The Quarterly Journal of Economics 134*(3), 1343–1404.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics 132*(2), 665–712.

Kremer, M., S. Gallant, O. Rostapshova, and M. Thomas (2021). Is Development Economics a Good Investment? Evidence on Scaling Rate and Social Returns from USAID's Innovation Fund. Technical report, United States Agency for International Development.

Lerner, J. (2000). The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program. *The Journal of Private Equity 3*(2), 55–78.

Lerner, J. (2012). *The Architecture of Innovation: The Economics of Creative Organizations*. Harvard Business Review Press.

Lerner, J., J. Liu, J. Moscona, and D. Yang (2024). Appropriate Entrepreneurship? The Rise of China and the Developing World. Working Paper 32193, National Bureau of Economic Research.

Lerner, J. and R. 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*(3), 237–261.

Lerner, J. and A. Seru (2022). The Use and Misuse of Patent Data: Issues for Finance and Beyond. *The Review of Financial Studies 35*(6), 2667–2704.

Mansfield, E. (1991). Academic Research and Industrial Innovation. *Research Policy 20*(1), 1–12.

Mansfield, E. (1995). Academic Research Underlying Industrial Innovations. *The Review of Economics and Statistics 77*(1), 55–65.

Mansfield, E. (1998). Academic Research and Industrial Innovation: An Update of Empirical Findings. *Research Policy 26*(7–8), 773–776.

Manso, G. (2011). Motivating Innovation. *The Journal of Finance 66*(5), 1823–1860.

Marx, M. and E. Scharfmann (2024). Does Patenting Promote the Progress of Science? Technical report.

Mayer, S. (2022). Financing breakthroughs under failure risk. *Journal of Financial Economics 144*(3), 807–848.

Moretti, E., C. Steinwender, and J. Van Reenen (2023). The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers. *The Review of Economics and Statistics*, 1–46.

Moscona, J. (2021). Flowers of Invention: Patent Protection and Productivity Growth in

US Agriculture. SSRN Scholarly Paper 3924439.

Moscona, J. and K. Sastry (2022). Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture. SSRN Scholarly Paper 3744951.

Myers, K. and W. Y. Tham (2023). Money, Time, and Grant Design. Working Paper 24-037, Harvard Business School.

Myers, K. R. and L. Lanahan (2022). Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy. *American Economic Review 112*(7), 2393–2423.

Nadauld, T. D., B. A. Sensoy, K. Vorkink, and M. S. Weisbach (2019). The liquidity cost of private equity investments: Evidence from secondary market transactions. *Journal of Financial Economics 132*(3), 158–181.

Nyarko, J. (2021). Stickiness and Incomplete Contracts. *University of Chicago Law Review 88*(1), 1–79.

Phang, J., Y. Zhao, and P. Liu (2023). Investigating Efficiently Extending Transformers for Long Input Summarization. In H. Bouamor, J. Pino, and K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 3946–3961.

Richman, B. (2011). Contracts Meet Henry Ford. *Hofstra Law Review 40*, 77–86.

Ritter, J. R. (2015). Growth Capital-Backed IPOs. *Financial Review 50*(4), 481–515.

Ruttan, V. W. (1959). Usher and Schumpeter on Invention, Innovation, and Technological Change. *The Quarterly Journal of Economics 73*(4), 596–606.

Sampat, B. and H. L. Williams (2019). How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome. *American Economic Review 109*(1), 203–236.

Stein, J. C. (2003). Agency, Information, and Corporate Investment. In G. Constantinides, R. Stulz, and M. Harris (Eds.), *Handbook of the Economics of Finance*, pp. 111–165.

Stein, J. C. (2005). Why Are Most Funds Open-Ended? *The Quarterly Journal of Economics 120*(1), 247–272.

Sternitzke, C. (2010). Knowledge Sources, Patent Protection, and Commercialization of Pharmaceutical Innovations. *Research Policy 39*(6), 810–821.

Summers, L. and E. Balls (2015). Report of the Commission on Inclusive Prosperity. Technical report, Center for American Progress.

Swensen, D. F. (2000). *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*. Simon and Schuster.

Terry, S. J. (2023). The Macro Impact of Short-Termism. *Econometrica 91*(5), 1881–1912.

Thomas, C. (2019). Experimentation with Reputation Concerns – Dynamic Signalling

with Changing Types. *Journal of Economic Theory 179*, 366–415.

Toole, A. A. (2012). The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry. *Research Policy 41*(1), 1–12.

Townsend, R. R. (2015). Propagation of Financial Shocks: The Case of Venture Capital. *Management Science 61*(11), 2782–2802.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997). University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology 5*(1), 19–50.

WIPO (2022). *Guide to the International Patent Classification (2022)*.

## Appendix


**A Data Construction** **A2**

A.1 Sample of Patents and Citations . . . . . A2

A.2 Statutory Requirements for Citing Prior Art . . . . . A2

A.3 Constructing the Descriptions of Cooperative Patent Classification . . . . . A3

A.4 Constructing Dataset with Venture Capital Funding . . . . . A6

A.5 Constructing Dataset with Federal Funding . . . . . A9

A.6 A Short Note on Matching . . . . . A16

A.7 Tagging Patents . . . . . A20

**B Gestation Lag Measure** **A23**

B.1 Comparison with Existing Literature . . . . . A23

**C Example of a Venture Capital Limited Partnership Agreement** **A24**

**D Contracts in Other Sources of Funding** **A27**

**E Other Tables and Figures** **A29**

# A. Data Construction

In this section, I describe each of the datasets that are used as input to construct the six analyses datasets.

## A.1. Sample of Patents and Citations

`g_us_patent_citation.tsv` contains data that links patents to all the other patents that cite them. The file downloaded from Pitchbook on November 15, 2023 contains 135,946,772 rows. It contains data on 8,537,802 patents. After dropping all patents that have non-numeric patent IDs (these include patents that begin with letters such as D, R, and P corresponding to Design, Re-issue, and Plant patents) or receive citations from patents that have non-numeric patent IDs, the dataset has information on follow-on citations received by 7,559,926 unique patent IDs. There are 119,283,105 patent-cited pairs. The last citing patent was filed on 2023-05-01.

I link my patent-level sample (described in Section 4) to the citation data. Of the 3,336,866 patents in my sample, 785,748 (23%) patents do not have a single follow-on citation. I further drop 6,219 rows if the citing patents were filed before 1980-01-01. There are 64,396,878 patent-cited by pairs in the data.

## A.2. Statutory Requirements for Citing Prior Art

The relevant legal requirement for citing patent art is addressed in 35 U.S. Code §102 - Conditions for patentability. To understand what constitutes prior art, I turn to the USPTO, which provides guidance to prospective patentees on how to search for prior art.[24] According to patent law in the U.S., an entity is not entitled to a patent if the claimed invention was: patented, described in a printed publication, in public use, on sale, or available to the public *before the effective filing date of the claimed invention.* The wide gamut of available knowledge that is considered "prior art" is further emphasized in USPTO's definition of prior art:

"Information known publicly before the effective filing date of a U.S. patent application is called prior art. Prior art consists of information available to the public including:

- U.S. patents and published patent applications.

[24]Source: https://www.uspto.gov/sites/default/files/documents/Basics-of-Prior-Art-Searching.pdf.

- Foreign patents and published patent applications
- Journal and magazine articles
- Books, manuals, and catalogs
- Websites
- Conference proceedings
- Scientific papers

Prior art may be published on paper or electronic format."

Since this definition does not allow any statutory limitation on what knowledge can be considered prior art, the boundaries determining what information needs to be cited (or what existing information can prevent the granting of a patent) come from 35 U.S. Code §301 - Citation of prior art and written statements. This section reads as follows:

"Any person at any time may cite to the Office in writing–

1. prior art consisting of patents or printed publications which that person believes to have a bearing on the patentability of any claim of a particular patent; or
2. statements of the patent owner filed in a proceeding before a Federal court or the Office in which the patent owner took a position on the scope of any claim of a particular patent."

This section allows anyone to write to the patent office with regards to a particular patent if they believe that *any* claim in the patent misses citation to prior art.

These two sections–the first on what constitutes novelty, and therefore permits patenting, and the second on who is allowed to address missing art in a patent–collectively emphasize that all relevant knowledge needs to be cited by a patent. There is no deadline to what needs to be cited, i.e., patents must cite previous patents even if they have expired.

## A.3. Constructing the Descriptions of Cooperative Patent Classification

In this section, I describe how I construct descriptions of CPC subclasses which I then use to assign various sources of funding to technologies. I use information from the Guide to the International Patent Classification (henceforth, the "IPC Guide") (WIPO, 2022) and

the Guide to the Combined Patent Classification (henceforth the "CPC Guide"). The guide to the CPC states, on page 5, that "Unless otherwise noted, CPC structure, rules and principles are identical to those of the IPC. Therefore, only the distinguishing features of the CPC with regard to the IPC will be described in this Guide."

CPC provides the schema and definitions in XML files that are available for bulk download[25] I use the XML files for both CPC schema and definitions. Files were last downloaded on 2024-01-05. These files are available at the level of the 4-digit CPC hierarchy, which is also called the "subclass." The 1-digit, 2-digit, and 3-digit hierarchies are called the sections, subsections, and classes respectively. The 4-digit subclasses are further divided into groups, and groups themselves are divided into main groups and subgroups.

To see an example of the content available in the bulk download, see the scheme for the subclass A01J at https://www.uspto.gov/web/patents/classification/cpc/html/cpc-A01J.html and its associated definition at https://www.uspto.gov/web/patents/classification/cpc/html/defA01J.html. I do not take into account the information that is conveyed via color of text and curly brackets on the website since these are usually markers of how the CPC differs from the IPC. Since I am using the 4-digit subclass, there is no distinction between the CPC and IPC at that level.

I construct the description of a subclass by combining its title, titles of all its constituent main groups, and the definition statements that correspond to each of these components. I do not include the titles of sections, subsections, and classes. I omit titles of subgroups that are not part of the files in the definitions folder since they need to be combined with titles of other hierarchies in complicated ways that are hard to automate within code (see paragraphs 25-29 in the IPC guide for details.)

The following excerpts from the IPC guide justify my methodology.

Paragraph 68 of the IPC guide states that the titles of the sections, subsections, and classes are "only broadly indicative of their content and do not define with precision the subject matter falling under the general indication of the title. ... By contrast, it is the intention in the Classification that the titles of subclasses, taking into consideration any references, definitions or notes associated therewith, define as precisely as possible the scope of the subject matter covered thereby. The titles of main groups and subgroups, again subject to any references, definitions or notes associated therewith, precisely define

[25]Source: https://www.cooperativepatentclassification.org/cpcSchemeAndDefinitions/bulk accessed in February, 2024.

the subject matter covered thereby."

Descriptions of subclasses and groups includes limiting and non-limiting "references," that delimit and contextualize the technologies. Similar technologies that are nevertheless categorized elsewhere are mentioned under Limiting References. Non-limiting references point the user to other subclasses that might be relevant for the user, or that which contain a technology's application (see paragraph 39 of (WIPO, 2022) for details on limiting and non-limiting references.) I do not include the keywords associated in either type of reference list.

Since my preferred level of hierarchy is the CPC subclass, I follow paragraph 69 that describes how to construct the scope of the subclass. It states the following:

"(a) The subclass title which describes, as precisely as is possible in a small number of words, the main characteristic of a portion of the whole body of knowledge covered by the Classification, this portion being the field of the subclass to which all its groups relate.

(b) Any limiting references which follow the subclass title or its class title. These references indicate certain parts of the field described by the title which are covered by other subclasses and are therefore excluded. These parts may constitute a substantial part of the field described by the title and, thus, the limiting references are in some respects as important as the title itself. For example, in subclass A47D – FURNITURE SPECIALLY ADAPTED FOR CHILDREN – a considerable part, namely school benches or desks, of the subject matter covered by the title is excluded in view of a reference to particular groups of subclass A47B, thus considerably altering the scope of subclass A47D.

(c) Any limiting references which appear in groups of a subclass and which refer subject matter to another class or subclass also restrict the scope of the subclass in question. For example, in subclass B43K – IMPLEMENTS FOR WRITING OR DRAWING – writing points for indicating or recording apparatus are referred out of group 1/00 to group 15/16 of subclass G01D, thereby reducing the scope of the subject matter covered by the title of subclass B43K.

(d) Any notes appearing under the subclass title or its class, subsection or section title. Such notes may define terms or expressions used in the title, or elsewhere, or clarify the relation between the subclass and other places.

The titles of subclasses sometimes do not embrace the titles of all main groups under them. However, the scope of a subclass should always be understood to include all subject matter specifically stated in the titles of its main groups.

70. A more detailed explanation of the scope of a subclass is provided by its definition,

where it is available. In particular, the exact boundaries of any subject matter added to the scope of the subclass by its main group titles are specified."

To see how patents are allocated to patents, I refer the reader to paragraphs 77 to 105 in the IPC guide.

**Notices of Change** A final note on the stability of the technology classification system. Like any classification system, the CPC system has made changes to both the schema and definitions. The CPC publishes notices of changes when the classification system is updated. The full list of notices of change can be found here: https://www.uspto.gov/web/patents/classification/cpc/html/cpc-notices-of-changes.html. Manual inspection of several randomly chosen notices suggests that the changes happen mostly at the subgroup level. This means the subclass itself (and its attendant definition and title) mostly remains unchanged.

**Final Dataset** My dataset has information on 643 CPC classes. On average, the description of a CPC class has 1,820 words with a standard deviation of 4,175 words. The maximum is 49,780 words and the minimum is 20 words. The median is 698 words (which is about a page).

Before constructing Term Frequency-Inverse Document Frequency vectors, I pre-process the CPC descriptions, which is standard practice and increases the accuracy of the Natural Language Processing Methods. First, I turn all text to lower case, then I lemmatize the words, and then remove punctuations and stop words. Lemmatizing converts words to their root or stem form, using their position in the sentence to contextualize their meaning. As an example for how lemmatization affects text in my context, consider the following sentence that shows up in the description of CPC subclass A61F: "Devices for processing exclusively digital data." Lemmatization converts this to the following: "Devices for process exclusively digital data." Lemmatization therefore reduces the dimensionality of the TF-IDF vectors and increases accuracy of matching.

## A.4. Constructing Dataset with Venture Capital Funding

I use the dataset covered under the license between Harvard University and PitchBook and released as of 2023-09-01. The subset of the dataset that is shared with Harvard

**Table A1:** Examples of Mapping Between PitchBook Sectors and CPC Subclasses

| PitchBook Sector | CPC | CPC Title | CPC Share |
|---|---|---|---|
| Information Technology | G06Q | Information and communication technology [ict] specially adapted for administrative, commercial, financial, managerial or supervisory purposes | 0.12 |
| | H04B | Transmission | 0.05 |
| | G06V | Image or video recognition or understanding | 0.04 |
| Healthcare | G16H | Healthcare informatics, ie information and communication technology [ict] specially adapted for the handling or processing of medical or healthcare data | 0.19 |
| | A61P | Specific therapeutic activity of chemical compounds or medicinal preparations | 0.09 |
| | A61M | Devices for introducing media into, or onto, the body | 0.04 |
| Consumer Products and Services (B2C) | A23L | Foods, foodstuffs, or non-alcoholic beverages | 0.06 |
| | H04W | Wireless communication networks | 0.04 |
| | G10G | Representation of music | 0.03 |
| Business Products and Services (B2B) | G06Q | Information and communication technology [ict] specially adapted for administrative, commercial, financial, managerial or supervisory purposes | 0.08 |
| | G09F | Displaying | 0.06 |
| | G06V | Image or video recognition or understanding | 0.03 |
| Financial Services | G06Q | Information and communication technology [ict] specially adapted for administrative, commercial, financial, managerial or supervisory purposes | 0.15 |
| | G07F | Coin-freed or like apparatus | 0.06 |
| | C30B | Single-crystal growth | 0.04 |
| Energy | F24S | Solar heat collectors | 0.20 |
| | G21H | Obtaining energy from radioactive sources | 0.06 |
| | G21D | Nuclear power plant | 0.06 |
| Materials and Resources | A23L | Foods, foodstuffs, or non-alcoholic beverages | 0.07 |
| | F23G | Cremation furnaces | 0.04 |
| | C02F | Treatment of water, waste water, sewage, or sludge | 0.03 |

*Notes:* The first column reports the sectors used by PitchBook to categorize its firms, ordered by the number of firms in each sector. For each sector, I report the three CPC subclasses that are most frequently linked to the firms in that sector using the NLP techniques discussed in Section 3.2. The columns titled CPC and CPC Title report the CPC subclass and its associated title.

**Table A2:** Total Patenting and Investment by PitchBook Industries

| Sector | Companies | Number of Investments | Value of Investments | Patents |
|---|---|---|---|---|
| Business Products and Services (B2B) | 7,088 | 29,517 | 221 | 22,386 |
| Consumer Products and Services (B2C) | 8,913 | 37,245 | 317 | 50,022 |
| Energy | 778 | 2,649 | 70 | 4,300 |
| Financial Services | 1,570 | 8,014 | 119 | 1,848 |
| Healthcare | 11,022 | 49,882 | 492 | 92,973 |
| Information Technology | 25,772 | 131,022 | 933 | 209,700 |
| Materials and Resources | 714 | 2,858 | 25 | 5,442 |
| Total | 55,857 | 261,187 | 2,176 | 386,671 |

*Notes:* A tabulation of investments and patents by PitchBook sectors.

includes firms that are in the "VC-backed universe" with no restriction on years (all years of data available to PitchBook is covered by the Harvard University license) or geography. PitchBook defines the universe variable as follows:

" Represents the category that a company falls into based on the type of investors it has or has had in the past. Each universe incorporates companies based on current or past ownership statuses. A company may fall into multiple universes based on its current and prior owners. (If a company was VC-backed and PE-backed at some point, their Company Universe be "VC-backed, PE backed.") "

As such, the PitchBook dataset covers firms that receive VC funding at least once.

To these set of firms, we observe all funding provided by other types of investors. PitchBook tags each investor's Primary Investor Type; there are a total of 37 types of investors. The 20 largest (by number of companies they have invested in) are Venture Capital, PE/Buyout, Angel (individual), Corporation, Accelerator/Incubator, PE-Backed Company, Corporate Venture Capital, Growth/Expansion, Angel Group, Government, Asset Manager, Family Office, Hedge Fund, Mezzanine, VC-Backed Company, Real Estate, Investment Bank, Impact Investing, Lender/ Debt Provider, and Sovereign Wealth Fund.

My sample of companies is those that receive a deal by an investor with Primary Investor Type of "Venture Capital" at least once in the first 20 years of their founding. These firms must have a headquarter in the U.S. I further restrict my deal-level sample to deals

made between 1980 and 2022 (inclusive.) This limits my sample to 56, 316 firms.

The exhibit, Table A18, shows a snapshot of the dataset for the venture capital firm called Kleiner Perkins.

**Data on exits** I combine two datasets to collect information on investor exits from companies: `DealSellerRelation.dat` and `InvestorExitRelation.dat`. The first dataset contains deal-level information on exits from companies: whether the exit was partial or full (i.e., whether the investor sold 100% of her share in the company), if partial, what percent of the company was sold, and the fund from which the exiter sold her shares. To this information, I add the second dataset which includes the type of exit that the deal refers to; the 5 most common types of exits are [Merger/Acquisition, Out of Business, Secondary Transaction - Private, IPO, and Buyout/LBO].

I merge exit-level data to investment-level data using information on the investor's ID, ID of the exiting fund (if available), and the company ID. In the case of multiple partial exits of a fund from a company, I keep the earliest exit from a fund. I illustrate why this is a reasonable choice with a specific example. The investor Fritsch exited from Tesla twice: once on 2010-06-29 and for the second time on 2014-10-22. The first exit is labeled as an IPO and the second exit is labeled as a Secondary Transaction - Open Market. In this situation, the IPO is the first publicly-observable valuation event, and therefore the exit relevant to my setting. Sometimes, an investor exits from a company in two separate deals on the same day. A frequent way this happens is if a company declares bankruptcy and is acquired in the same day. Or, there can be two separate secondary transactions. I do not judge which one is the "correct" exit type to keep in such situations and arbitrarily choose the exit that starts with the alphabetically smaller character.

Definitions of variables in the PitchBook dataset available online at `https://pitchbook.com/news/pitchbook-report-methodologies`, which was downloaded on 4/22/2024.

## A.5. Constructing Dataset with Federal Funding

Here, I describe award-level datasets for 6 government agencies: the Department of Defense (DOD), Department of Health and Human Services (HHS), Department of Energy (DOE), the National Science Foundation (NSF), and the National Aeronautics and Space Administration (NASA).

**HHS (NIH)** The National Institutes of Health, which is a major source of funding for health innovation, is housed under the HHS.[26]

I download bulk export files from https://reporter.nih.gov/exporter/projects and https://reporter.nih.gov/exporter/abstracts. These links contain one folder for each fiscal year; each folder contains bibliographic and abstract data respectively for each grant awarded in that year. This data were downloaded in January, 2024.

The NIH data starts in 1985. I transform the dataset for each year so that I have total costs for each unique project, where a project is identified by its unique title and abstract. I identify a unique project by the grant number, which is stored in the FULL PROJECT NUM variable. I extract a project's unique identifier by eliminating the first digit, which contains the application type, and all the digits that come after the hyphen, which contain the suffix. [27] I calculate the total costs for a project by year by summing up all costs of its sub-components in a given year. A project might show up in my dataset in multiple years if the project raised NIH grants over multiple years.

For example, one of the unique project IDs I create is F32AI007244. This unique project is linked to the following grant numbers: 1F32AI007244-01 (awarded in 1985), 1F32AI007244-01X1 (awarded in 1985), 3F32AI007244-01S1 (awarded in 1985), 3F32AI007244-01S2 (awarded in 1986), 5F32AI007244-02 (awarded in 1986), 3F32AI007244-02S1 (awarded in 1986), 3F32AI007244-01S3 (awarded in 1986), 5F32AI007244-03 (awarded in 1987), and 3F32AI007244-02S2 (awarded in 1987.) I add up the sums for each year, and assign the total dollar amount to the year in which it was granted. A different way to do this would be to associate all dollars granted to the first year of the grant. I take the former route for two reasons. First, it allows me to deflate the grant amounts correctly. Second, this approach parallels how I allocate venture capital funding. A company can raise deals in multiple years; I do not assign all deals to the first year that a deal was raised.

For a handful of observations, there seem to be two different grant numbers but holding the same information, for example N01CM047673-000 and N01CM047673-001. I do not treat these differently; since these two rows share the same constructed project ID, I sum the total cost associated with to get the year annual for the project number 01CM047673.

The codebook for NIH is available at the following url: https://api.reporter.nih.gov/documentsData%20Elements%20for%20RePORTER%20Project%20API_

[26] Source: https://www.nih.gov/about-nih/who-we-are.

[27] Information on the NIH grant ID can be found on the following link: https://www.med.upenn.edu/orss/assets/user-content/documents/nih-numbering-2.pdf.

V2.pdf. I benchmark my constructed dataset against the annual totals presented by the Congressional Research Service on the National Institutes of Health (NIH) Funding: FY1996-FY2024 accessed at https://sgp.fas.org/crs/misc/R43341.pdf. My annual totals fall very slightly short of the ones reported in Table 1 for the years 1996-2022.

The funding data is only available when funding is made by NIH, CDC, and the FDA.

Table A3 reports, for each year, the total number of grants, the number of grants that were missing information on the total cost, the number of grants that were missing information on abstracts, and the sum of NIH grants in that year (in current dollars.)

**NSF** Data for awards from the NSF comes from https://www.nsf.gov/awardsearch/download.jsp. This data were downloaded in January, 2024.

Table A4 summarizes the NSF dataset.

**SBIR/STTR** SBIR/STTR award data is available for bulk download at https://www.sbir.gov/sbirsearch/award/all. The codebook for this dataset is available online at https://www.sbir.gov/dictionary/data-dictionary-award. This data were downloaded in January, 2024.

The raw bulk file with SBIR/STTR grant data contains information on 156,755 unique values of Agency Tracking Number, which is a unique code assigned by the agency/branch that the grant is submitted to for phase 1. Since SBIR/STTR awards are granted in two phases, I identify each round of funding provided by the SBIR/STTR program by combining, for each grant, the Agency Tracking Number, the Agency that is making the award, the Phase of the grant, and the year of the award. There are 201,185 total unique awards based on the new ID variable corresponding to 202,337 rows in the data. I drop 345 observations for being exact duplicates on this constructed ID variable. I drop a further 27 observations for missing information on the Award Amount. A further 124 rows show up as having duplicated award IDs. Most of these rows correspond to two different award amounts. From inspecting the SBIR/STTR codebook and website, it is unclear whether these correspond to two different amounts (and thus should be added together to get the total) or if one row is providing a corrected version of the other. In the absence of this information, the best I can do is to take the average value of the two amounts. This removes a further 62 rows from the data. Finally, I drop 729 additional rows because in these cases, two rows have the same ID and Award Amount information but different abstracts. It is unclear which one to keep, so I sort the dataset alphabetically

**Table A3:** Summary of NIH Grant Data

| Year | Number of Awards | Total Current Dollars | Missing Abstract | Missing Amount |
|---|---|---|---|---|
| 1985 | 35,265 | 4,455,552,757 | 0.27 | 0.14 |
| 1986 | 35,572 | 4,644,518,846 | 0.27 | 0.14 |
| 1987 | 38,006 | 5,495,674,742 | 0.30 | 0.14 |
| 1988 | 38,437 | 5,961,222,260 | 0.28 | 0.13 |
| 1989 | 38,066 | 6,626,621,954 | 0.30 | 0.09 |
| 1990 | 39,933 | 7,008,030,389 | 0.27 | 0.13 |
| 1991 | 42,336 | 7,639,139,142 | 0.26 | 0.14 |
| 1992 | 42,172 | 8,197,077,683 | 0.23 | 0.14 |
| 1993 | 41,036 | 8,382,990,254 | 0.21 | 0.13 |
| 1994 | 42,687 | 8,990,265,282 | 0.24 | 0.14 |
| 1995 | 42,379 | 9,258,831,134 | 0.29 | 0.14 |
| 1996 | 41,997 | 9,751,493,964 | 0.87 | 0.12 |
| 1997 | 43,629 | 10,390,891,581 | 0.18 | 0.12 |
| 1998 | 43,617 | 11,162,779,805 | 0.11 | 0.08 |
| 1999 | 45,887 | 12,845,833,256 | 0.10 | 0.07 |
| 2000 | 48,256 | 14,558,871,515 | 0.09 | 0.08 |
| 2001 | 49,695 | 15,670,175,973 | 0.08 | 0.07 |
| 2002 | 53,134 | 17,693,664,156 | 0.07 | 0.08 |
| 2003 | 56,156 | 19,789,369,265 | 0.07 | 0.08 |
| 2004 | 58,415 | 20,592,918,511 | 0.09 | 0.09 |
| 2005 | 59,116 | 21,249,189,741 | 0.08 | 0.10 |
| 2006 | 58,778 | 21,003,750,215 | 0.06 | 0.10 |
| 2007 | 62,388 | 26,462,616,942 | 0.10 | 0.09 |
| 2008 | 62,948 | 26,117,037,383 | 0.11 | 0.10 |
| 2009 | 68,673 | 37,853,763,470 | 0.08 | 0.05 |
| 2010 | 69,089 | 38,268,899,866 | 0.09 | 0.05 |
| 2011 | 62,368 | 32,502,232,216 | 0.09 | 0.06 |
| 2012 | 62,474 | 32,900,627,250 | 0.10 | 0.06 |
| 2013 | 60,441 | 31,416,124,669 | 0.09 | 0.06 |
| 2014 | 60,057 | 32,471,230,372 | 0.09 | 0.07 |
| 2015 | 57,217 | 32,660,361,008 | 0.04 | 0.03 |
| 2016 | 57,150 | 31,920,281,324 | 0.02 | 0.03 |
| 2017 | 57,384 | 31,527,020,658 | 0.00 | 0.03 |
| 2018 | 59,837 | 33,681,167,424 | 0.00 | 0.03 |
| 2019 | 62,311 | 36,293,275,228 | 0.00 | 0.03 |
| 2020 | 64,174 | 40,489,232,449 | 0.06 | 0.03 |
| 2021 | 64,993 | 42,002,786,842 | 0.01 | 0.03 |
| 2022 | 66,549 | 43,112,559,976 | 0.00 | 0.03 |

*Notes*:This table reports annual statistics for the award-level data from this federal agency. Dollar amounts are in current dollars. Please see Appendix A.5 for details about how this dataset was constructed. A high fraction of grants in 1996 are missing an abstract. There is no clear reason why; the raw abstract file itself is half of the size of the file for the previous and following years.

**Table A4:** Summary of NSF Grant Data

| Year | Number of Awards | Total Current Dollars | Missing Title | Missing Abstract |
|---|---|---|---|---|
| 1980 | 8,151 | 1,276,575,582 | 0.00 | 1.00 |
| 1981 | 7,118 | 840,325,729 | 0.00 | 1.00 |
| 1982 | 6,514 | 1,348,864,410 | 0.00 | 1.00 |
| 1983 | 7,401 | 1,543,386,457 | 0.00 | 1.00 |
| 1984 | 7,893 | 1,686,852,503 | 0.00 | 0.95 |
| 1985 | 8,036 | 1,381,069,143 | 0.00 | 0.88 |
| 1986 | 7,758 | 1,404,649,729 | 0.00 | 0.73 |
| 1987 | 8,073 | 1,955,120,251 | 0.00 | 0.12 |
| 1988 | 8,703 | 1,905,843,757 | 0.00 | 0.06 |
| 1989 | 9,062 | 3,691,459,064 | 0.00 | 0.06 |
| 1990 | 10,082 | 2,092,140,616 | 0.00 | 0.08 |
| 1991 | 10,732 | 2,376,362,971 | 0.00 | 0.08 |
| 1992 | 10,627 | 3,185,712,392 | 0.00 | 0.08 |
| 1993 | 10,060 | 3,199,629,008 | 0.00 | 0.06 |
| 1994 | 9,956 | 3,361,891,755 | 0.00 | 0.05 |
| 1995 | 9,771 | 2,834,415,458 | 0.00 | 0.09 |
| 1996 | 9,447 | 3,221,961,958 | 0.00 | 0.06 |
| 1997 | 10,257 | 3,688,904,413 | 0.00 | 0.07 |
| 1998 | 9,599 | 3,410,807,751 | 0.00 | 0.08 |
| 1999 | 10,404 | 3,346,170,702 | 0.00 | 0.06 |
| 2000 | 10,370 | 5,925,045,497 | 0.00 | 0.05 |
| 2001 | 9,848 | 4,610,809,887 | 0.00 | 0.06 |
| 2002 | 10,955 | 6,599,834,685 | 0.00 | 0.04 |
| 2003 | 11,556 | 5,973,354,686 | 0.00 | 0.01 |
| 2004 | 10,721 | 5,851,604,748 | 0.00 | 0.01 |
| 2005 | 10,730 | 5,140,753,365 | 0.00 | 0.01 |
| 2006 | 11,063 | 5,420,296,528 | 0.00 | 0.01 |
| 2007 | 12,223 | 5,643,010,271 | 0.00 | 0.01 |
| 2008 | 12,612 | 7,155,602,405 | 0.00 | 0.01 |
| 2009 | 15,430 | 8,415,344,940 | 0.00 | 0.01 |
| 2010 | 13,094 | 7,133,414,817 | 0.00 | 0.03 |
| 2011 | 11,652 | 6,390,738,794 | 0.00 | 0.03 |
| 2012 | 12,165 | 7,050,523,710 | 0.00 | 0.01 |
| 2013 | 11,502 | 5,929,065,306 | 0.00 | 0.01 |
| 2014 | 12,109 | 6,056,284,040 | 0.00 | 0.01 |
| 2015 | 13,072 | 6,272,777,653 | 0.00 | 0.01 |
| 2016 | 12,594 | 7,113,094,816 | 0.00 | 0.01 |
| 2017 | 12,289 | 8,162,671,071 | 0.00 | 0.01 |
| 2018 | 12,235 | 7,017,957,020 | 0.00 | 0.01 |
| 2019 | 13,043 | 7,405,076,867 | 0.00 | 0.01 |
| 2020 | 13,300 | 6,969,767,431 | 0.00 | 0.01 |
| 2021 | 12,651 | 7,196,264,405 | 0.00 | 0.01 |
| 2022 | 11,458 | 5,615,968,653 | 0.00 | 0.01 |

*Notes:* This table reports annual statistics for the award-level data from this federal agency. Dollar amounts are in current dollars. Please see Appendix A.5 for details about how this dataset was constructed.

on abstract within each award ID and keep just the first. The final dataset has 201,174 unique IDs. I also update the grant awards by adding title and abstract together.

There are two dates associated with each SBIR/STTR grant: Proposal Award Date and Award Year. The former is the first day of contract performance, while the latter describes the fiscal year of the award.[28] I assign the award amount to the Award Year in keeping with how I clean the other federal grant data.

Table A5 summarizes the SBIR/STTR dataset.

Table A6 splits the aggregate SBIR/STTR funding by agency.

**DOD** I download abstract-level data for grants made by the Department of Defense from the Defense Technical Information Center (DTIC) last accessed at https://dtic.dimensions.ai/discover/grant on May 8, 2024. The DTIC operates under the Secretary of Defense and is tasked with preserving and disseminating research. More information about the DTIC can be found at https://discover.dtic.mil/about/. On its website, it reports "descriptive abstracts of public, unclassified DoD-allocated grant awards and information." The data is not available in an easily downloadable form; I access the data through its collaboration with the Dimensions platform that makes DoD documents available to the public. As of May 10, 2024, the Dimensions platform reported 141,407 DoD grant awards. I scrape information on the following variables for these grants: Grant Number, Title, Research Organization, Abstract, Funding Amount, Funding Period, and Research Categories. The scraped dataset has 118,543 observations. The missing grants were not accessible to the scraping query for various reasons. The data spans 1983 to 2024. The DTIC website reports that this interface contains "contains a collection of publicly-searchable DoD research and non-research grants awarded since December 2014 – containing detailed project abstracts and insight into DoD funding allocations" but the scraped data goes as far back as 1983.

Table A7 summarizes the DOD dataset.

**DOE** The grants allocation at the Department of Energy is conducted through its Office of Science. List of grant identifiers for winners is available at https://science.osti.gov/Universities/sc-in-your-state/. Digitized files exist for years 2012-2023; PDFs are provided for the years 2006-2011. I digitize this set of files. Files contain, for each fiscal year, all the awards granted by the Office of Science, Department of Energy.

[28] Both NIH and this data are at the fiscal year-level. I assume these are calendar years as elsewhere.

**Table A5:** Summary of SBIR/STTR Grant Data

| Year | Number of Awards | Total Current Dollars | Missing Abstract |
|---|---|---|---|
| 1983 | 785 | 37,849,979 | 0.00 |
| 1984 | 1,286 | 148,955,308 | 0.00 |
| 1985 | 1,822 | 193,140,774 | 0.00 |
| 1986 | 2,152 | 302,535,708 | 0.00 |
| 1987 | 2,741 | 268,177,799 | 0.00 |
| 1988 | 2,639 | 389,523,245 | 0.00 |
| 1989 | 2,781 | 393,463,322 | 0.00 |
| 1990 | 3,189 | 447,566,980 | 0.00 |
| 1991 | 3,363 | 449,176,821 | 0.00 |
| 1992 | 3,405 | 517,216,900 | 0.00 |
| 1993 | 3,959 | 622,947,814 | 0.00 |
| 1994 | 3,985 | 595,805,899 | 0.00 |
| 1995 | 4,322 | 948,045,891 | 0.00 |
| 1996 | 4,020 | 914,709,862 | 0.00 |
| 1997 | 4,741 | 1,122,731,707 | 0.00 |
| 1998 | 4,446 | 1,058,204,577 | 0.00 |
| 1999 | 4,945 | 1,163,480,738 | 0.00 |
| 2000 | 5,622 | 1,110,499,194 | 0.00 |
| 2001 | 6,150 | 1,226,509,574 | 0.00 |
| 2002 | 7,165 | 1,579,748,701 | 0.00 |
| 2003 | 7,389 | 1,843,147,113 | 0.00 |
| 2004 | 7,371 | 2,411,353,018 | 0.00 |
| 2005 | 6,938 | 2,300,907,756 | 0.00 |
| 2006 | 6,640 | 2,448,703,598 | 0.00 |
| 2007 | 6,294 | 1,936,992,890 | 0.00 |
| 2008 | 6,203 | 2,338,688,652 | 0.00 |
| 2009 | 6,608 | 2,270,626,135 | 0.00 |
| 2010 | 7,189 | 2,594,393,725 | 0.00 |
| 2011 | 6,094 | 2,309,064,975 | 0.00 |
| 2012 | 5,630 | 2,256,208,744 | 0.00 |
| 2013 | 5,087 | 2,088,650,056 | 0.00 |
| 2014 | 5,246 | 2,294,582,050 | 0.00 |
| 2015 | 5,167 | 2,498,056,674 | 0.00 |
| 2016 | 5,358 | 2,661,815,110 | 0.00 |
| 2017 | 6,016 | 3,172,982,634 | 0.00 |
| 2018 | 5,657 | 2,844,341,878 | 0.00 |
| 2019 | 7,034 | 3,799,150,171 | 0.00 |
| 2020 | 7,295 | 3,889,026,655 | 0.00 |
| 2021 | 6,872 | 3,607,762,980 | 0.00 |
| 2022 | 6,577 | 3,727,737,870 | 0.00 |
| 2023 | 991 | 408,781,681 | 0.00 |

*Notes:* This table reports annual statistics for the award-level data from this federal agency. Dollar amounts are in current dollars. Please see Appendix A.5 for details about how this dataset was constructed.

**Table A6:** Summary of SBIR/STTR Grant Data by Agency

| Agency | Number of Awards | Total Current Dollars |
|---|---|---|
| Department of Agriculture | 3,852 | 594,325,573 |
| Department of Commerce | 1,740 | 259,043,225 |
| Department of Defense | 96,015 | 32,809,314,180 |
| Department of Education | 1,384 | 251,994,632 |
| Department of Energy | 15,378 | 5,176,967,234 |
| Department of Health and Human Services | 45,396 | 19,385,157,901 |
| Department of Homeland Security | 1,128 | 345,809,860 |
| Department of Interior | 15 | 485,280 |
| Department of Transportation | 1,226 | 252,956,459 |
| Environmental Protection Agency | 1,958 | 214,818,942 |
| National Aeronautics and Space Administration | 18,134 | 4,474,003,470 |
| National Science Foundation | 14,779 | 3,413,619,787 |
| Nuclear Regulatory Commission | 168 | 14,699,172 |

*Notes:* The value of Total Current Dollars is calculated by adding up award amounts across multiple years. See Appendix A.5 for details on this dataset.

For the awards, I observe at least their ID, the awardee institution, and the total funding amount.

The DOE also provides public abstracts for its grants but the abstracts need to be accessed via the PAMS Public Abstract Search for one grant at a time using the award ID in the state-level files (Link to the PAMS portal: https://www.energy.gov/science/portfolio-analysis-and-management-system-pams.) However, abstract data is not available before 2014. I therefore only use title-level data for the Department of Energy.

As I do for other datasets, I allocate the funding granted in each year to that year's total, even if it was part of a grant that was awarded earlier, or if the funding was meant to be used for subsequent years.

Table A8 summarizes the DOE dataset.

## A.6. A Short Note on Matching

The objective of the matching exercise is to assign each firm and grant to a technology class based on the similarity of their descriptions. It would be too costly to do this manually since we have to link about 50,000 firms and several million grants to about 600 technology classes. We instead use an algorithmic approach that capitalizes on recent advances in Natural Language Processing. The exact choice of technique depends on the setting and the classic need to trade between Type I and Type II errors, which are dubbed

**Table A7:** Summary of DOD Grant Data

| Year | Number of Awards | Total Current Dollars | Missing Abstract |
|---|---|---|---|
| 1983 | 282 | 15,337,145 | 0.00 |
| 1984 | 460 | 49,384,868 | 0.00 |
| 1985 | 718 | 94,187,229 | 0.00 |
| 1986 | 1,181 | 156,787,174 | 0.00 |
| 1987 | 1,461 | 139,512,918 | 0.00 |
| 1988 | 1,296 | 196,920,741 | 0.00 |
| 1989 | 1,273 | 190,507,402 | 0.00 |
| 1990 | 1,550 | 215,666,474 | 0.00 |
| 1991 | 1,573 | 198,692,280 | 0.00 |
| 1992 | 1,510 | 254,065,114 | 0.00 |
| 1993 | 2,217 | 530,067,487 | 0.00 |
| 1994 | 1,799 | 263,632,700 | 0.00 |
| 1995 | 2,164 | 591,597,671 | 0.00 |
| 1996 | 2,283 | 559,147,741 | 0.00 |
| 1997 | 2,568 | 713,044,734 | 0.00 |
| 1998 | 2,427 | 715,068,902 | 0.00 |
| 1999 | 2,153 | 324,366,310 | 0.00 |
| 2000 | 1,206 | 591,476,775 | 0.00 |
| 2001 | 2,929 | 975,326,965 | 0.00 |
| 2002 | 3,368 | 991,605,218 | 0.00 |
| 2003 | 3,505 | 1,328,437,064 | 0.00 |
| 2004 | 3,255 | 1,378,835,071 | 0.00 |
| 2005 | 3,930 | 1,679,488,331 | 0.00 |
| 2006 | 4,076 | 1,702,859,170 | 0.00 |
| 2007 | 3,480 | 1,430,401,910 | 0.00 |
| 2008 | 3,845 | 1,881,930,543 | 0.00 |
| 2009 | 3,795 | 1,790,538,198 | 0.00 |
| 2010 | 3,625 | 2,038,724,160 | 0.00 |
| 2011 | 3,711 | 1,946,816,478 | 0.00 |
| 2012 | 3,582 | 1,740,610,260 | 0.00 |
| 2013 | 3,122 | 1,604,592,605 | 0.00 |
| 2014 | 2,943 | 1,647,844,223 | 0.00 |
| 2015 | 4,274 | 2,835,188,713 | 0.00 |
| 2016 | 4,500 | 4,031,137,550 | 0.00 |
| 2017 | 4,714 | 3,762,881,670 | 0.00 |
| 2018 | 4,320 | 3,559,406,663 | 0.00 |
| 2019 | 4,239 | 4,017,722,588 | 0.00 |
| 2020 | 4,855 | 5,029,842,114 | 0.00 |
| 2021 | 4,575 | 7,359,211,000 | 0.00 |
| 2022 | 4,759 | 4,358,817,897 | 0.00 |
| 2023 | 4,889 | 4,694,773,956 | 0.00 |

*Notes:* This table reports annual statistics for the award-level data from this federal agency. Dollar amounts are in current dollars. Please see Appendix A.5 for details about how this dataset was constructed.

**Table A8:** Summary of DOE Grant Data

| Year | Number of Awards | Total Current Dollars | Missing Title |
|---|---|---|---|
| 2006 | 2,418 | 634,869,000 | 0.00 |
| 2007 | 2,382 | 572,739,000 | 0.00 |
| 2008 | 4,021 | 873,845,000 | 0.00 |
| 2009 | 2,479 | 1,064,885,000 | 0.00 |
| 2010 | 2,629 | 1,048,059,000 | 0.00 |
| 2011 | 2,673 | 879,222,000 | 0.00 |
| 2012 | 2,683 | 845,464,000 | 0.00 |
| 2013 | 2,617 | 778,068,000 | 0.00 |
| 2014 | 2,604 | 899,772,978 | 0.00 |
| 2015 | 2,571 | 913,952,561 | 0.00 |
| 2016 | 2,404 | 941,660,801 | 0.00 |
| 2017 | 2,417 | 897,829,577 | 0.00 |
| 2018 | 2,550 | 1,072,767,261 | 0.00 |
| 2019 | 2,616 | 1,004,791,279 | 0.00 |
| 2020 | 2,733 | 1,054,951,493 | 0.00 |
| 2021 | 2,909 | 1,059,797,107 | 0.00 |
| 2022 | 3,046 | 1,194,379,233 | 0.00 |

*Notes:* This table reports annual statistics for the award-level data from this federal agency. Dollar amounts are in current dollars. Please see Appendix A.5 for details about how this dataset was constructed.

"precision" and "recall" problems in the field of machine learning.

Section 3.2 has already described how I construct a vocabulary and Inverse Document Frequency weights using the descriptions of CPC subclasses. To do this most accurately, I limit the technology descriptions to the first 100 words, which is approximately the length of the descriptions of firms and abstracts of grants. Implementing the matching technique using the full, unabridged versions of technology descriptions caused errors (based on a manual evaluation of the matching outcome.) In practice, limiting the description to the first 100 words limits it to the title of the main CPC subclass and the most related top 2 or 3 groups within that subclass.

One draw-back of the TF-IDF approach using unigrams is that certain words can get a large weight (since they are rare) which drives mis-matching. For example, the presence of the word "pipeline" in any venture capital-backed firm's description caused it to be allocated to F17D, which is a CPC subclass covering natural gas pipelines. I reviewed several dozen matches to collect key words with such a characteristic and removed them from descriptions of firms and abstracts of grants before linking to CPC subclasses. These words are: pipelines, operate, provide, platform, develop, target, application, life-saving, data, therapy, cloud. I further amended two words: "web" to "internet", and "block chain" to "digital currency." The word "web" was causing firms to be linked to package-

technology which uses webs of paper. The latter was being linked to technologies to make metal chains.

In keeping with best practice, I eliminate all matches that have a cosine value of less than 0.05 (Kelly et al., 2021).

An alternative, more systematic method to eliminate single-word-led matches is to use bigrams rather than unigrams. However, the length of the firm descriptions and abstracts is too short for this approach; using bigrams caused matching rates to plummet. I only use bigrams to match CPC subclasses to NIH data, since the NIH dataset provides long descriptions.

I used two other methods to check if they would give me better accuracy. The first was to compute sentence embeddings using the BERT transformer. The BERT transformer has been used in the finance literature to map firms, using their 10Ks, to industries (Hoberg and Phillips, 2016). The alleged advantage of using transformers over bean-counting approaches like the TF-IDF is that transformers are supposed to understand the meaning of each word based on its location in the sentence. The vector for each paragraph is the average ("mean-pooled") vector of each word in the paragraph, where the vector for each word is created based on the context that the transformer has learned from being trained on billions of documents. However, this method did not work well in my setting. The transformer is not trained for this specific task, and so the vector representation of each technology description was quite similar to the other. In other words, the transformer was not properly over-weighting key words. One reason for this failure in my setting is the well-known understanding that transformers are unable to handle long inputs (Phang et al., 2023).

I also implemented the VoyageAI reranker, which is an example of a cross-encoder. A cross-encoder does not produce a sentence embedding; instead, it processes a pair of paragraphs simultaneously and outputs a value between 0 and 1 indicating the similarity of the pair. (Within this framing, the TF-IDF is an example of a bi-encoder since it produces paragraph-level vectors, after which we compute a measure of similarity.) This approach performed like BERT because it was missing the specific context of the task. TF-IDF worked best by highly weighting the technical key words that were relevant for a technology.

## A.7. Tagging Patents

My primary method to allocate funding to technologies is, as described elsewhere, by using the similarity between text-based descriptions of technologies and funding. I also use an alternative method, where I compare the patents funded by each funding source directly. In this section, I describe how I link patents to three sources of funding: venture capital, federal agencies, and public firms.

**Patents assigned to venture capital firms** I identify patents filed by venture capital-backed assignees by linking the names of assignees in the patents dataset to names of ventures provides by PitchBook. The linkage between patents and any firm-level dataset is complicated by the fact that patent offices do not require firms to submit names in any standardized format. As a result, the same firm may appear in several forms: IBM, for example, appears as International Business Machine, Inter. Bus. Mach, and so on. See Autor et al. (2020) for additional exposition on the difficulties of matching firm and patent datasets.

I use a machine learning package designed for record linkage, with a specific focus on linking datasets on the names of firms (Arora and Dell, 2024). The advantage of using a machine learning algorithm in this exercise is that the package is already trained on a large corpus of similar tasks and can therefore make reasonable guesses that two entries are the same.

I first prepare the PitchBook dataset. For each firm, PitchBook reports a constant Company ID and up to 4 names: CompanyName, CompanyFormerName, CompanyAlsoKnownAs, and CompanyLegalName. Since firms may change names or appear sometimes as acronyms, I use all four names corresponding to each Company ID. I process these further by converting all names to lower case and removing all periods and commas from the firm names. The PitchBook dataset released in 2023-09-01 contains 526,506 unique Company IDs. I restrict this dataset to firms that are headquartered in the U.S. There are 214,659 such firms with 480,932 unique names. There are 509,266 unique company ID-company name pairs in the dataset restricted to firms headquartered in the U.S.

Since the record linkage occurs using names of firms, I cannot keep two entries with the same name. In the case of duplicated names, I keep the entry corresponding to the larger firm (based on number of employees as reported in the 2023-09-01 release.) I drop 321 firms at this step. After this, the dataset has 799 names that show up twice (affecting 1,598 company name-company ID pairs.) I drop the rows where the company name is

the 2nd, 3rd, or 4th name corresponding to a company ID. This further eliminates 4 company IDs. The final PitchBook dataset on companies has 214,334 company IDs, 480,449 company names. From the PatentsView data, I use the file g-assignee-disambiguated.tsv, linked to g-location-disambiguated.tsv. The dataset g-assignee-disambiguated.tsv contains data on 475,277 unique assignee IDs. I link the assignee dataset to the location dataset to limit my sample to firms located in the U.S. Note that PatentsView assigns several assignees to multiple locations. I do not attempt to determine which location assignment is correct. I only keep assignees for whom at least one location is listed as the U.S. There are 236,879 unique assignee IDs who are located in the U.S.

I use the following command to link the PatentsView assignee dataset to PitchBook:

```
df_matched = lt.merge(patentsview_assignees,
pitchbook_assignees, merge_type='m:1',
on='assignee_name', model='all-MiniLM-L6-v2')
```

The command allows multiple assignees in the PatentsView dataset to be linked to the same assignee in PitchBook. It uses a model from Hugging Face, all-MiniLM-L6-v2, which maps sentences and paragraphs to a 384-dimensional dense vector space and computes similarities between sentence pairs. This dataset has been fine-tuned on a dataset containing 1 billion sentence pairs.[29]

Visual inspection suggests that the matching algorithm does reasonably well in linking the records. For example, it determines that the assignee *cidra corporate services inc* in the patents dataset has a similarity score of 0.875 with *cidera inc* in the PitchBook dataset. Other simpler matches, such as *teledyne limited* and *teledyne* get a higher score of 0.90. I keep only those pairs that are matched with a confidence above 0.865, which corresponds to the 75th percentile of confidence scores in this matching exercise. The choice of this threshold is ultimately arbitrary and attempts to balance Type I and Type II errors: the probability of matching two assignee names even when they are not the same, and the probability of missing a match when they are most likely the same firm. 59,519 unique assignee IDs in the PatentsView dataset are linked to a PitchBook assignee with a high level of confidence. They are linked to 45,263 unique Company IDs from the PitchBook dataset.

I further subset this dataset to include only those PitchBook firms that are VC-backed (as described in the previous subsection.) There are 21,429 such assignees.

[29]For more details on this package, go to https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2.

**Patents that received federal funding** I use the Government Patent Register constructed by (Gross and Sampat, 2024) to identify patents that were funded by federal agencies. I use the variables that start with *agency* to identify patents that received funding from the corresponding agency. Note that one patent may be assigned to several federal agencies.

I restrict the sample to patents that were granted between 1980 and 2022 and where the variable "Any Government Interest" is 1. There are 176,884 such patents. In the patent-level sample, 168,890 patents are assigned to the federal government in my sample. My sample is missing 7,995 patents from the Government Patent Register because 6,023 patents are missing information on either the assignee or the assignee's location. I drop a further 1,971 patents because none of their assignees are from the U.D.

**Table A9:** Patenting by Different Government Agencies in the Government Patent Register

| Agency | Patent Count | Fraction | CumFrac |
|---|---|---|---|
| Defense | 57,454 | 0.34 | 0.34 |
| Health and Human Services | 49,650 | 0.29 | 0.63 |
| Energy | 30,015 | 0.18 | 0.80 |
| National Science Foundation | 16,187 | 0.09 | 0.90 |
| National Aeronautics and Space Administration | 8,551 | 0.05 | 0.95 |
| Agriculture | 3,495 | 0.02 | 0.97 |
| Commerce | 2,505 | 0.01 | 0.98 |
| Veterans Affairs | 1,037 | 0.01 | 0.99 |
| Environmental Protection Agency | 537 | 0.00 | 0.99 |
| Interior | 507 | 0.00 | 1.00 |
| Transportation | 447 | 0.00 | 1.00 |
| Justice | 164 | 0.00 | 1.00 |
| Treasury | 97 | 0.00 | 1.00 |
| Office of Scientific Research and Development | 0 | 0.00 | 1.00 |
| Total | 170,646 | 1.00 | 1.00 |

*Notes:* This table summarizes the dataset built by Gross and Sampat (2024). I use the dummy variables *agency* to count the number of patents assigned to each agency. A patent may be assigned to multiple agencies in this table.

**Patents assigned to public firms** To identify patents allocated to firms in Compustat–which covers the universe of public firms in the U.S.–I use the dataset constructed and used by Dyevre (2023). There are 2,612,445 unique patent ids (where the first digit is non-string) allocated to Compustat firms between 1980 and 2023 in this dataset. In my patent data, there are 1,681,508 patent IDs that are assigned to Compustat firms. I drop 4,585 patents from the Dyevre (2023) because they were either not Utility patents or did not have a number as the first digit in their ID. I drop 7,041 patents because they were either missing assignee or location information. Finally, I drop 919,311 patents because none of their assignees were in the U.S.

# B. Gestation Lag Measure

## B.1. Comparison with Existing Literature

The canonical measure of gestation lags comes from a series of papers by Mansfield (1991, 1995, 1998), who surveyed managers of large firms and asked them about the recency of the scientific innovations that their product development was based on. While a direct comparison of the estimates is difficult because the papers use different definitions of industries and technologies, the gestation lags constructed using half-lives in my data resemble those reported those in these papers. Mansfield defines gestation lags as "the average time interval between first commercial introduction of a new product or process based on a recent academic research finding and the finding itself." He surveys managers at several firms to construct this data and finds measures that range between 2.4 years and 10.3 years. This is similar to range in my preferred measure that uses the half-life of patent citations of 2 to 13 years. Furthermore, Mansfield finds that Information processing has shorter lags, between 2.4 to 5.2 years, relative to drugs and medical products, that have the longest gestation lags between 8 to 10.3 years. In my preferred measure of gestation lags, surgery and medical equipment subclasses also correspond to the largest gestation lags.

I provide a brief overview of the other measures of innovation-related lags proposed in the literature. The most related paper is Sternitzke (2010) that also uses patent filing and citations to calculate time to commercialization. He estimates an average lag of 11 years for patents related to drugs. Budish et al. (2015) use the expected survival time for an organ-stage-specific cancer as their relevant measure. Drugs designed to target cancers

that have longer survival times need to consequently be in clinical trials for longer, which shortens the length of protection that a fixed-term, 20-year patent can provide them. Bastianin et al. (2022) study effect of becoming a CERN supplier on patenting and use the time between the start of the supplier relationship and patenting as their measure of gestation lag. They argue that this time estimates the time to build a commercial product after learning specifics and know-how from CERN. Hall et al. (1986) use the distributed lag between research and development expenditures and patent production to measure the delay until commercialization. Toole (2012) measures the lags between investment in public research and application to the FDA. He estimates that it takes approximately 17 years between public investment in basic biomedical research and the first related patent.

**Reliance on Science** I use a new dataset released by (Marx and Scharfmann, 2024) that constructs patent-paper-pairs. In the dataset, they link patents to the academic papers on which they build using information on the names of authors and content of the patent and papers. They also calculate the number of dates between when the academic paper was published and the patent was filed. Each patent-paper pair is given a score of 1, 2, 3 or 4 based on the authors' confidence of the quality of the match, where 4 is the highest confidence; I drop the the pairs corresponding to a score of 1. I merge this dataset onto the patent spine constructed for this study so I ensure that I calculate the basicness based only on patents relevant to my study.

I use this dataset in two ways to construct a technology's reliance on science. First, I calculate the number of patents in a CPC subclass that rely on any academic article. Second, for the patents that rely on a paper, I calculate the average delay between when the paper was published and when the patent was filed.

Other papers have measured basicness of patents slightly differently such as the generality of follow-on citations across different technology classes (Jaffe et al., 1998) or overall share of its own citations going to scientific papers (Trajtenberg et al., 1997).

# C. Example of a Venture Capital Limited Partnership Agreement

I include the first two pages of a model Limited Partnership Agreement made available by the Institutional Limited Partners Association, accessed from https://ilpa.org/wp-content/uploads/2020/07/ILPA-Model-LPA-Term-Sheet-WOF-Version.pdf on 07-07-2024.

**The ILPA Model Limited Partnership Agreement (Whole of Fund Version)**

**Term Sheet Summary**

*This term sheet is a summary of the principal terms of the Model LPA (issued in October 2019 by ILPA and as further amended in July 2020) and should be read in conjunction with the Model LPA which shall apply unless expressly varied by this term sheet.*

*This term sheet is for information purposes only, and legal advice should be obtained before implementing the Model LPA or any of its terms. The terms of this term sheet are not exhaustive and are not intended to be legally binding and are subject to the agreement and signing by all relevant parties of a detailed and legally binding limited partnership agreement.*

*Capitalised and undefined terms shall bear the meaning ascribed to them in the Model LPA.*

| | |
|---|---|
| **Fund** | [NAME OF FUND], a [NAME OF FUND JURISDICTION]-registered limited partnership. Parallel vehicles controlled by the General Partner may be formed to co-invest with the Fund on substantially the same terms.<br><br>Other investment vehicles may be also established in order to meet the specific needs of investors and/or to address specific tax, legal, regulatory or similar considerations. |
| **General Partner** | [NAME OF GENERAL PARTNER], a [FORM OF ENTITY], registered in [NAME OF JURISDICTION]. |
| **Fund Manager** | [NAME OF MANAGER], a [FORM OF ENTITY], registered in [NAME OF JURISDICTION]. |
| **Maximum Size** | The aggregate Commitments to the Fund and any parallel vehicles shall not exceed [CURRENCY/AMOUNT]. |
| **Currency** | The currency of the Fund will be [CURRENCY]. |
| **General Partner's Commitment** | The General Partner and its Affiliates shall make and maintain an aggregate Commitment by subscribing for an Interest equal to at least [•]% of the aggregate Commitments of the Limited Partners. |
| **Initial Closing Date** | The initial closing will occur on the date determined by the General Partner when the aggregate Commitments equal or exceed [NUMBER]. |
| **Final Closing Date** | The Fund shall cease accepting Commitments on the date that is 12 months from the Initial Closing Date. |
| **Subsequent Closings** | Additional closings may occur at the discretion of the General Partner. Investors admitted after the Initial Closing Date will be treated as if they had invested in the Fund at the Initial Closing Date and will participate in all of the Fund's investments and will pay their pro rata share of all drawdowns made prior to their admission plus additional payment on such amounts (other than any drawdowns attributable to the management fee) at the rate of [•]% per annum |

(subject to appropriate adjustments for investments realised prior to their respective admission to the Fund).

**Term of the Fund**

[10] years from the Initial Closing Date, subject to two one-year extensions by the General Partner and with the prior consent of the Advisory Committee to the first extension and with the prior consent of a Majority in Interest to the second extension.

**Commitment Period**

The period from the Initial Closing Date to the [fifth] anniversary of the Initial Closing Date and subject to a one-year extension by the General Partner with the prior consent of the Advisory Committee.

The Commitment Period may be terminated earlier if all Commitments have been drawn down and invested or used to create Reserves or for the follow-on investments, or by a notice approved by [75]% in Interest.

**Investment Policy**

The objective of the Fund is to provide attractive, long-term investment return by investing in a diversified portfolio of [INVESTMENT POLICY].

**Investment Restrictions**

The Fund shall only make investments if they are consistent with the Investment Policy.

Without the prior written consent of the Advisory Committee, the Fund will not invest:

(1) [more than [•]% [or [•]% including the aggregate of any Capital Contributions for any Bridge Investments] of the aggregate Commitments in any single portfolio company];

(2) [more than [•]% [or [•]% including the aggregate of any Capital Contributions for any Bridge Investments] of the aggregate Commitments in any one industry sector];

(3) [more than [•]%[or [•]% including the aggregate of any Capital Contributions for any Bridge Investments] of the aggregate Commitments in any group of Portfolio Companies that are Affiliates];

(4) [in Portfolio Companies outside of the Target Region];

(5) [in a Follow-On Investment [18 months] after the end of the Commitment Period]/[more than [15]% of aggregate Commitments in Follow-On Investments after the end of the Commitment Period;]

(6) [in publicly listed securities (save for those in connection with the public offering of an existing Portfolio Company)];

(7) [in speculative investments];

# D. Contracts in Other Sources of Funding

Federal government uses a range of instruments to provide funding to organizations. I get information on instruments through the websites that provide award-level data.

usaspending.gov lists the following Award Types: Contracts, Contract DIVs, Grants, Direct Payments, Loans, and Other. These are further divided into the following sub-categories:

- Contracts
  - Blanket Purchase Agreements (BPA) Calls
  - Purchase Order (PO)
  - Delivery Orders (DO)
  - Definitive Contracts
- Contract DIV*
  - Government-Wide Acquisition Contract (GWAC)
  - Indefinite Delivery/ Requirements Contract
  - Indefinite Delivery/ Indefinite Quantity (IDIQ) Contract
  - Indefinite Delivery/ Definite Quantity Contract
  - Federal Supply Schedule (FFS)
  - Basic Ordering Agreement (BOA)
  - Blanket Purchase Agreements (BPA)
- Grants
  - Block Grant
  - Formula Grant
  - Project Grant
  - Cooperative Agreement
- Direct Payments
  - Direct Payment with Unrestricted Use

  - Direct Payment for Specific Use

- Loans
  - Direct Loans
  - Guaranteed/ Insured Loans

- Other
  - Insurance
  - Other Financial Assistance
  - Not Specified

*The Contract DIV refers to a federal grant instrument introduced in 2010 by Dr. Michael Kremer and USAID's Chief Innovation Officer, Dr. Maura O'Neill. The Development Innovation Venture (DIV) award type "provides funding through fixed-amount awards. These awards are designed to pay for results through milestone payments, rather than just reimbursing costs like traditional grants." Source: https://www.usaid.gov/div.

The website that posts government funding opportunities, grants.gov/search-grants, has four types of funding instrument types: Cooperative Agreements, Grant, Other, and Procurement Contract.

# E. Other Tables and Figures

**Table A10:** Summary of Innovation Funding

| Source | Total Amount (Billions) | Num. Funding Rounds/Awards | Num. Companies |
|---|---|---|---|
| Venture Capital | 2,168 | 107,016 | 55,709 |
| Compustat | 19,659 | - | 15,804 |
| SBIR/STTR | 90 | 192,859 | - |
| NSF | 302 | 451,026 | - |
| DOD | 82 | 110,927 | - |
| DOE | 19 | 44,459 | - |
| NIH | 992 | 460,448 | - |

*Notes:* Table reports the total funding amount allocated from each innovation funding source in my sample. Total Amount is reported in billions of constant 2022 dollars. Num. Funding Rounds/Awards reports the number of funding rounds in case of venture capital and the total number of awards granted by each agency in the case of SBIR/STTR, NSF, DOD, DOE, and NIH. Abstracts for DOE awards are only available starting 2006, so the DOE data is only available from 2006-2022. Appendix A.5 contains information on how I collect data on funding from each source.

**Table A11:** Patent Sample

**Panel A:** Patent-Level Summary Statistics

| Sample | Publications | VC | Gov | Compustat |
|---|---|---|---|---|
| Patents published between 1980-2022 | 8,087,644 | - | - | - |
| Utility patents | 7,301,319 | - | - | - |
| Patents not missing assignees | 6,614,994 | - | - | - |
| Patents not missing assignee and assignee location | 6,576,806 | - | - | - |
| Patents where at least 1 assignee is in the U.S. | 3,336,866 | 387,178 | 168,888 | 1,681,488 |

**Panel B:** Assignee-Level Summary Statistics

| Sample | Publications | Assignees | Assignee-Publications |
|---|---|---|---|
| All | 3,336,866 | 235,946 | 3,414,498 |
| VC | 387,178 | 19,809 | 387,695 |
| Gov | 168,888 | 10,799 | 182,706 |
| Compustat | 1,681,488 | 19,809 | 1,712,625 |

**Panel C:** Citations Summary Statistics

| Sample | Publications |
|---|---|
| All | 3,336,866 |
| Citations==0 | 785,748 |
| Citations>0 | 2,551,138 |

*Notes:* The earliest and latest date of issue in the patent sample are 1980-01-01 and 2022-12-27, respectively. I drop 771,132 non-utility patents, 15,114 patents that are withdrawn, 13 patents that start with R (Re-issue), and 66 patents that do not have a CPC code in the `g_cpc_current.tsv` file from PatentsView. I count a patent as being VC-backed (or backed by the government or a Compustat firm) if any of the assignees of a patent are tagged as that. See Section 4 for details on the construction of this sample.

**Table A12:** Examples of Technologies and Gestation Lags

| Quartile | CPC | Title | Num. Patents | Half-Life |
|---|---|---|---|---|
| Q1 | H04L | Transmission of digital information, eg telegraphic communication | 120,381 | 2.20 |
| Q1 | G06F | Electric digital data processing | 236,331 | 2.67 |
| Q2 | E21B | Earth or rock drilling | 29,106 | 6.79 |
| Q2 | H01R | Electrically-conductive connections | 23,848 | 6.96 |
| Q3 | A61M | Devices for introducing media into, or onto, the body | 27,072 | 7.25 |
| Q3 | B01D | Separation; evaporating; use of electrical or wave energy | 20,566 | 7.48 |
| Q4 | B65D | Containers for storage or transport of articles or materials | 23,537 | 8.17 |
| Q4 | H01H | Electric switches; relays; selectors | 9,646 | 8.43 |

*Notes:* This table presents the gestation lags of top two technologies (by the number of patents) in each quartile of gestation lags of technologies. The gestation lag in this table is measured using patent-level half lives. The Title column reports the a short snippet of the the full description created for each technology. For further details on the choice of the CPC hierarchical level and the construction of its description, please see Section 3.2.

**Table A13:** Summary Statistics of Gestation Lag Measures

| Gestation Lag | Mean | Std. Dev. |
|---|---|---|
| Half Life | 7.09 | 1.86 |
| Mean Age of Cites | 7.72 | 1.94 |
| Time to 10 | 10.10 | 3.28 |
| Time to 10, pre 2018 | 10.17 | 3.27 |

*Notes:* Table presents summary statistics of the four different measures of gestation lag as described in Section 4.

**Table A14:** Correlation Between Gestation Lag Measures

| | Half Life | Mean Age of Cites | Time to 10 | Time to 10, pre 2018 |
|---|---|---|---|---|
| Half Life | 1.00 | 0.99 | 0.82 | 0.81 |
| Mean Age of Cites | 0.99 | 1.00 | 0.83 | 0.82 |
| Time to 10 | 0.82 | 0.83 | 1.00 | 1.00 |
| Time to 10, pre 2018 | 0.81 | 0.82 | 1.00 | 1.00 |

*Notes:* Table presents pairwise correlation between measures of gestation lag as described in Section 4.

**Table A15:** Changes in CPC-Level Gestation Lags As a Function of Venture Capital Investment

| | CPC-Level Slope |
|---|---|
| Intercept | -0.410*** |
| | (0.004) |
| Deal Size (Billion) | 0.000 |
| | (0.000) |
| N | 584 |
| R2 | 0.00 |

*Notes:* This table presents estimates from a regression of a CPC-level slope measuring the decline of its half-life between 1980 and 2022 on the total dollar amount (in constant 2022 dollars) it received from the venture capital in that time.

**Table A16:** Correlates of the Half-Life Measure

| | Patent Quality | Patent-Level Half-Life | CPC-Level Half-Life |
|---|---|---|---|
| 0 | Half-Life | 1.00 | 1.00 |
| 1 | Total Cites, Adj. Grant Year | 0.13 | -0.38 |
| 2 | KPST Score, 1 Year | 0.02 | -0.31 |
| 3 | KPST Score, 5 Year | -0.02 | -0.62 |
| 4 | Marx-Feugi Delay from Science | -0.02 | 0.00 |
| 5 | Any Academic Paper | NaN | -0.21 |

*Notes:* This table reports correlations (at the patent- and CPC-level) of different measures of patent quality to patent's half-life of citations. These measures are discussed in Section 5.

**Table A17:** Summary of Fund-Level Venture Capital Data

**Panel A:** Fund-Level Summary Statistics

| Variable | Obs. | Mean | Std. Dev. | Median |
|---|---|---|---|---|
| Total companies | 3,508 | 17.73 | 24.71 | 9.00 |
| Total deals | 3,508 | 24.90 | 34.25 | 13.00 |
| Fund age at first investment | 3,508 | 0.00 | 0.00 | 0.00 |
| Fund age at all investments | 3,508 | 3.15 | 2.83 | 2.69 |

**Panel B:** Company-Level Summary Statistics

| Variable | Obs. | Mean | Std. Dev. | Median |
|---|---|---|---|---|
| Age at first deal | 24,775 | 2.99 | 3.45 | 2.00 |
| Total deals | 24,775 | 3.53 | 4.00 | 2.00 |
| Number of investors | 24,775 | 2.14 | 1.66 | 1.00 |
| File any patent | 24,775 | 0.38 | 0.49 | 0.00 |
| Age at first granted patent | 9,495 | 3.62 | 8.91 | 5.00 |
| Total number of patents | 9,495 | 33.30 | 364.55 | 5.00 |

**Panel C:** Investment-Level Summary Statistics

| Variable | Obs. | Mean | Std. Dev. | Median |
|---|---|---|---|---|
| Company age (in years) | 62,205 | 3.71 | 3.59 | 3.00 |
| Young company dummy | 62,205 | 0.78 | 0.42 | 1.00 |
| Previous investments | 62,205 | 0.87 | 1.29 | 0.00 |
| Fund age (in years) | 62,205 | 3.87 | 3.81 | 3.00 |
| First-time fund dummy | 62,205 | 0.19 | 0.39 | 0.00 |
| Missing deal size | 62,205 | 0.07 | 0.26 | 0.00 |
| Mean investment value (2022 dollars) | 57,760 | 8.24 | 48.98 | 3.58 |

*Notes:* PitchBook reports a unique identification key for each time a company raises funds, "funding round." Since multiple investors may participate in each funding round, I define an investment as the unique combination of an investor, investor fund, and funding round. The sample is restricted to first-time investments from a fund into a company. Section 6 provides further information about sample construction. Panel A summarizes funds' characteristics. Panel B summarizes characteristics of ventures when they receive investments from venture capital investors. Panel C summarizes characteristics of investments.

**Table A18:** Example of Investment Over the Fund's Lifecycle

| Vintage | KPCB II 1980 | KPCB III 1982 | KPCB II Annex Fund 1985 | KPCB IV 1986 | KPCB V 1989 | KPCB VI 1992 | KPCB VII 1994 |
|---|---|---|---|---|---|---|---|
| 1980 | 2 | - | - | - | - | - | - |
| 1981 | 2 | - | - | - | - | - | - |
| 1982 | - | 1 | - | - | - | - | - |
| 1985 | - | - | 2 | - | - | - | - |
| 1986 | - | - | - | 3 | - | - | - |
| 1987 | 1 | - | - | - | - | - | - |
| 1989 | - | - | - | 2 | 5 | - | - |
| 1990 | - | - | - | - | 1 | - | - |
| 1991 | - | 1 | - | - | 1 | - | - |
| 1992 | - | - | - | 1 | 1 | 1 | - |
| 1993 | - | - | - | - | 1 | 8 | - |
| 1994 | - | - | - | 2 | - | 2 | 7 |
| 1995 | - | - | - | - | - | - | 11 |
| 1996 | - | - | - | - | - | - | 2 |
| 1997 | - | - | - | - | - | 1 | - |
| 1998 | - | - | - | - | - | - | 1 |

*Notes:* This exhibit presents a snapshot of the funding allocation by a venture capital firm, Kleiner Perkins. Please see Section 6 for a discussion.

**Table A19:** Institutional Types of Limited Partners

| LP Type | Detailed LP Type | Fund Count |
|---|---|---|
| Pension | Corporate Pension | 1497 |
| Pension | Union Pension Fund | 428 |
| Pension | Government Agency | 355 |
| Pension | Public Pension Fund | 1370 |
| Endowment | Endowment | 764 |
| Endowment | Family Office (Multi) | 76 |
| Endowment | Family Office (Single) | 81 |
| Endowment | Foundation | 1726 |
| Other | Banking Institution | 189 |
| Other | Sovereign Wealth Fund | 216 |
| Other | Secondary LP | 87 |
| Other | Real Estate Investment Company | 26 |
| Other | Private Investment Fund | 78 |
| Other | Other Limited Partner | 34 |
| Other | High-net-worth investor | 123 |
| Other | Investment Advisor | 192 |
| Other | Insurance Company | 668 |
| Other | University (Non-Endowment) | 58 |
| Other | Fund of Funds | 1086 |
| Other | Economic Development Agency | 137 |
| Other | Direct Investment | 363 |
| Other | Corporation | 415 |
| Other | Money Management Firm | 260 |
| Other | Wealth Management Firm | 195 |

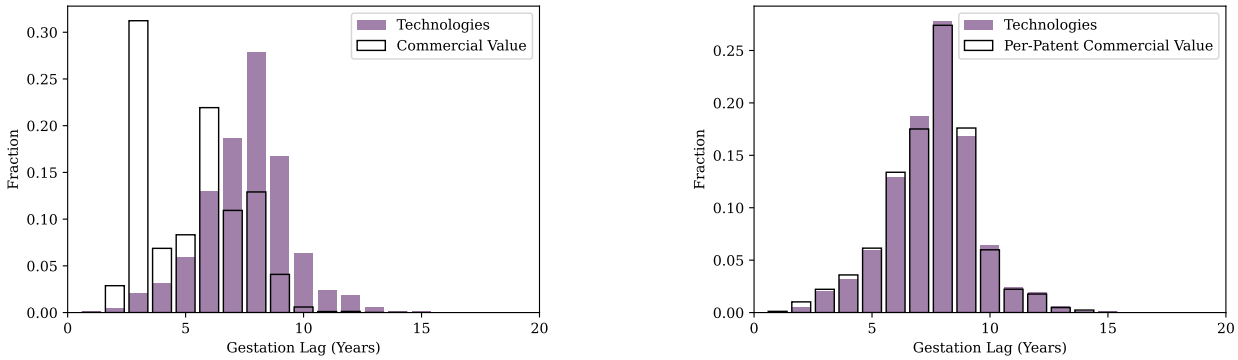
*Notes:* First column reports the coarse institutional type of Limited Partner that I use for the ensuing analyses. Column 2 reports the institutional types reported by PitchBook that are assigned to each coarse group. The last column reports the number of funds that each type of Limited Partner invests in between 1980 and 2022. Since multiple Limited Partners can invest in a fund, the column total exceeds the total number of funds in the data.

**Table A20:** Using Endowment LPs as IV

| | (1) Time to Exit | (2) Num Funds | (3) Time to Exit | (4) IV |
|---|---|---|---|---|
| Num Funds | -0.10*** | | | -0.14 |
| | (0.04) | | | (0.16) |
| 1=Endowment, 0=Pension | | 1.44*** | -0.20 | |
| | | (0.53) | (0.23) | |
| Intercept | 6.17*** | 3.07*** | 5.88*** | 6.31*** |
| | (0.20) | (0.20) | (0.12) | (0.56) |
| R-squared | 0.04 | 0.07 | 0.01 | 0.03 |
| R-squared Adj. | 0.03 | 0.06 | -0.00 | 0.03 |
| Observations | 134 | 134 | 134 | 134 |

*Notes:* Each column reports estimates from a separate regression. In Column (1), I regress the average time to exit for ventures that receive investment from venture capital firms between 2000 and 2007 on the number of funds that these firms raise in this time. The age of the venture is calculated at the time of the first investment from the firm into the venture. In column (2), I regress the number of funds on the treatment indicator, where the indicator is 1 if the venture capital firm raised capital from an endowment in 1999 and 0 if it raised capital from a pension fund. If a venture capital firm raised capital from both types of institutions, I assign them to treatment, i.e., that they raised capital from endowments. Column (3) reports the reduced form regression of time to exit on treatment status. Column (4) reports estimates from a two stage least squares regression where I instrument for the number of funds that a venture capital firm raised between 2000 and 2007 by the institutional type of limited partner it raised capital from in 1990. Robust standard errors are reported in parentheses.

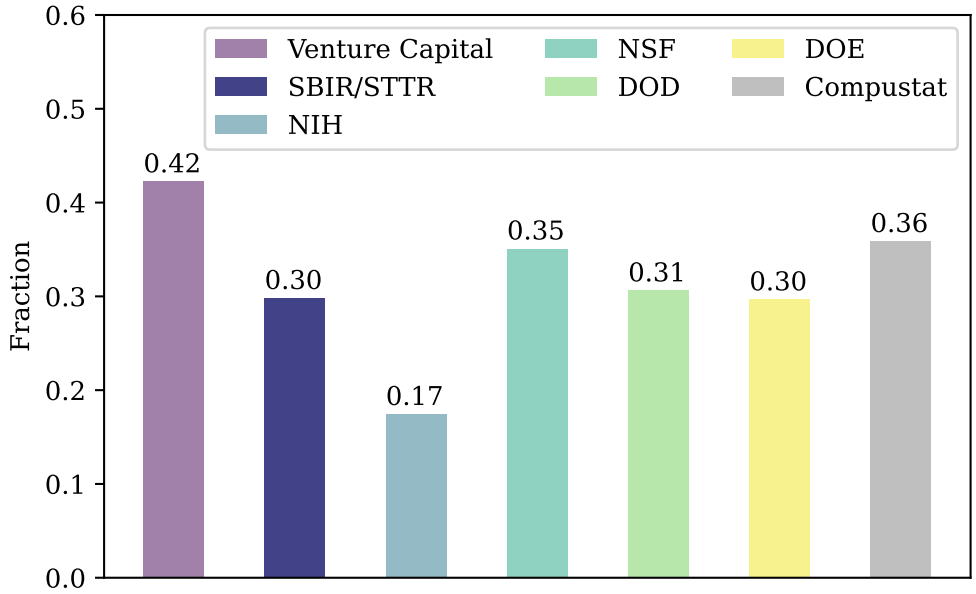
**Figure A1:** Distribution of Commercial Value




**(a)** Total Commercial Value **(b)** Per-Patent Commercial Value

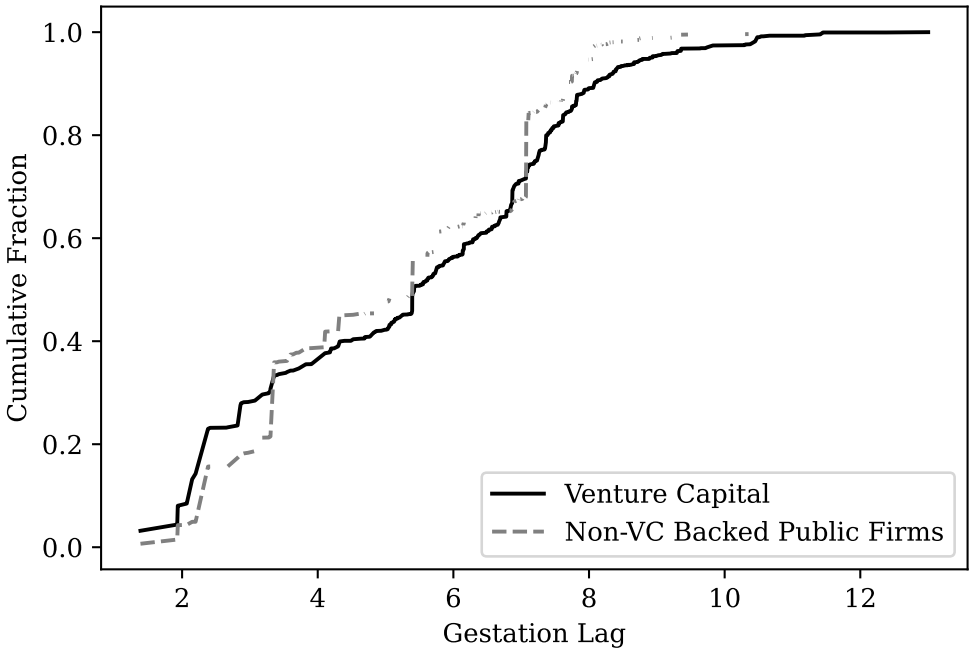
*Notes:* To construct these figures, I calculate the sum of patent-level real commercial value, as measured by Kogan et al. (2017), for all patents in a 4-digit subclass. The nominal patent-level value is converted to constant 2022 dollars using CPI values from FRED. The number of patents and their sum is calculated for the intersection of patents in the data from Kogan et al. (2017) and the spine of patents used in this paper as discussed in Section 3. The left panel plots the share of commercial value of patents belonging to each bin of gestation lags. The right panel plots the normalized commercial value, where the sum of all commercial value in a subclass is divided by the number of patents in it. For details on how gestation lag is constructed see Section 4.

**Figure A2:** Fraction of Funding Allocated to Technologies with Half-Life <= 5 Years



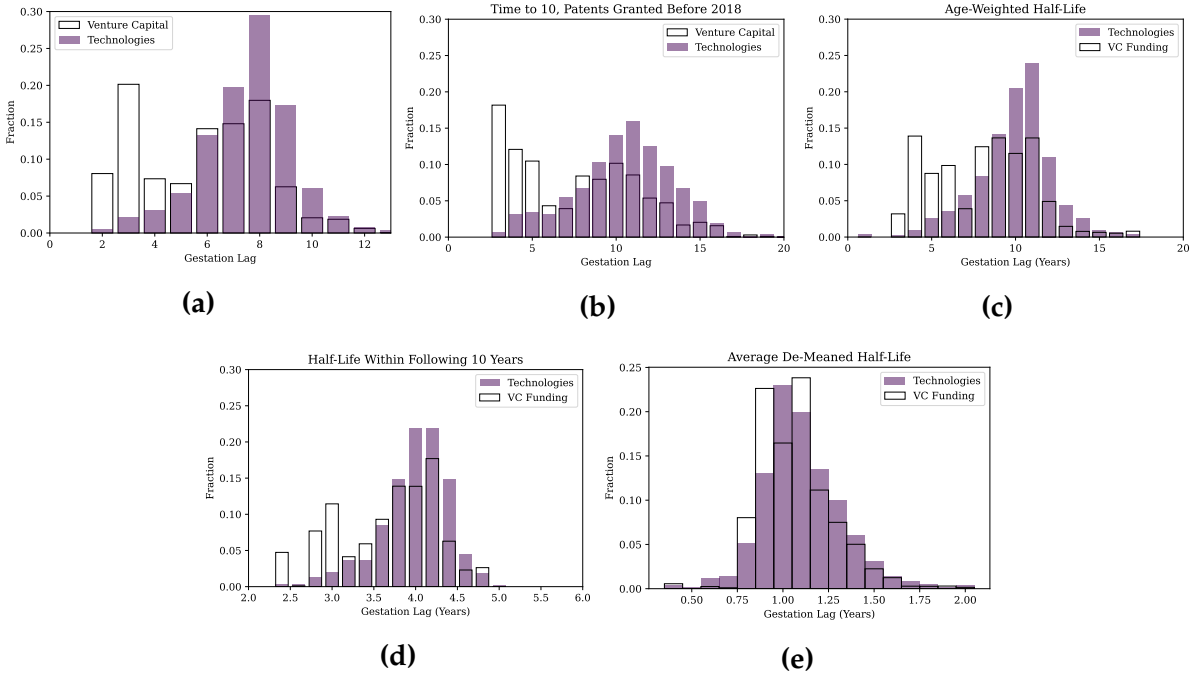

*Notes:* Each bar reports the fraction of the funding by each source allocated to technologies with gestation lags shorter than 5.42 years, which corresponds to the median gestation lag of the technologies funded by venture capital. See section Section 5 for a discussion of the corresponding cumulative distribution functions of these innovation funding sources.

**Figure A3:** Comparing the Allocation of Venture Capital



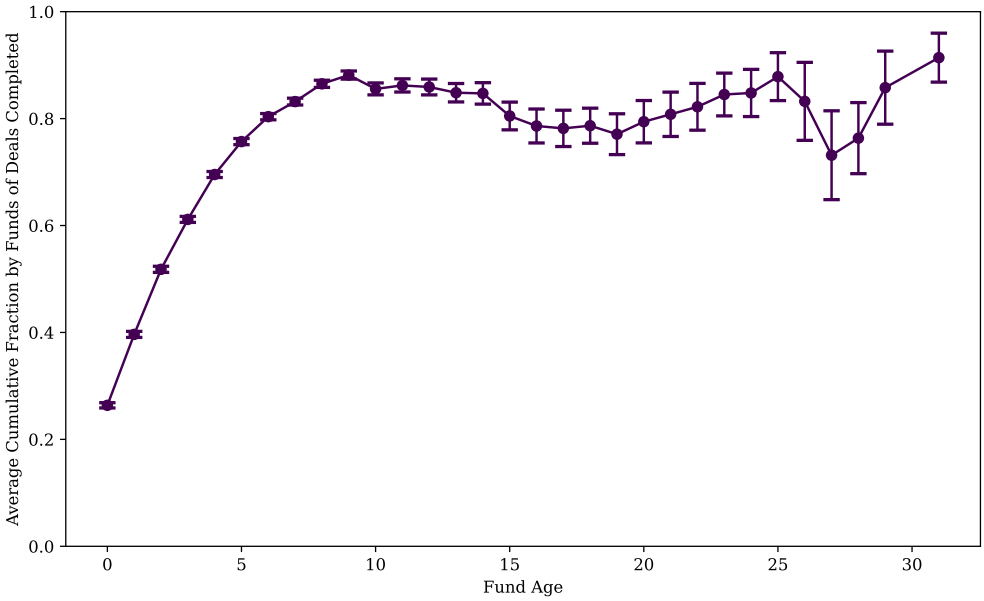

*Notes:* Figure plots the cumulative distribution of venture capital (in solid black) and Non-venture capital-backed public firms (in dashed grey). I identify public firms that were initially backed by venture capital firms using the dataset in Ritter (2015).

**Figure A4:** Comparing Alternative Measures of Gestation Lags



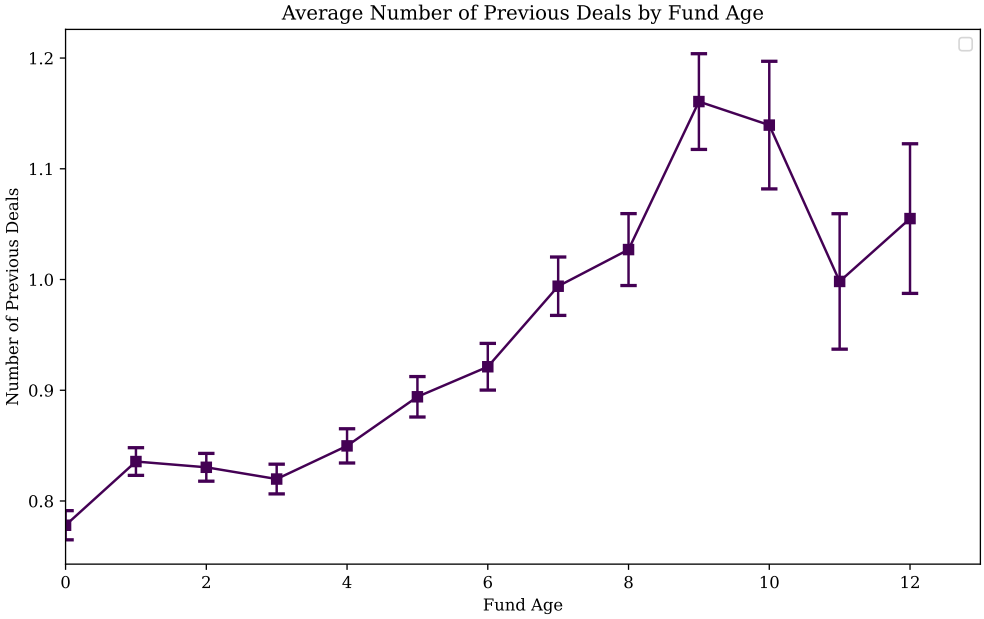

*Notes:* Panels present alternative measures of gestation lags (in purple bars), plotted alongside the distribution of venture capital (in black bars). I bin both funding amounts (in 2022 constant dollars) and the technologies so the bin at 3, for example, shows the fraction of technologies that have a gestation lag in the interval $(2,3]$ and the fraction of total venture capital funding that is assigned to them. Panel (a) presents the half-life measure that is used in the figures in the main text. Patent-level half-life is measured as the number of years between the date of publication of a patent when half of all its follow-on patents are filed. Panel (b) calculates the time until when a patent receives its first 10 citations. In Panel (c), I weigh each half-life with the weight of the patent, so that older patents have a higher weight when calculating subclass-level averages. Panel (b) measures half-life after restricting citations received in the following 10 years. To construct this figure, I keep patents granted on or before 2013. Last panel, panel (d), calculates subclass-level averages of half-lives after demeaning them using the mean half-life of all patents granted in the same year. For further details, please see Section 5.

**Figure A5:** Term Limits in the Data (Average of Fund-level Fractions)



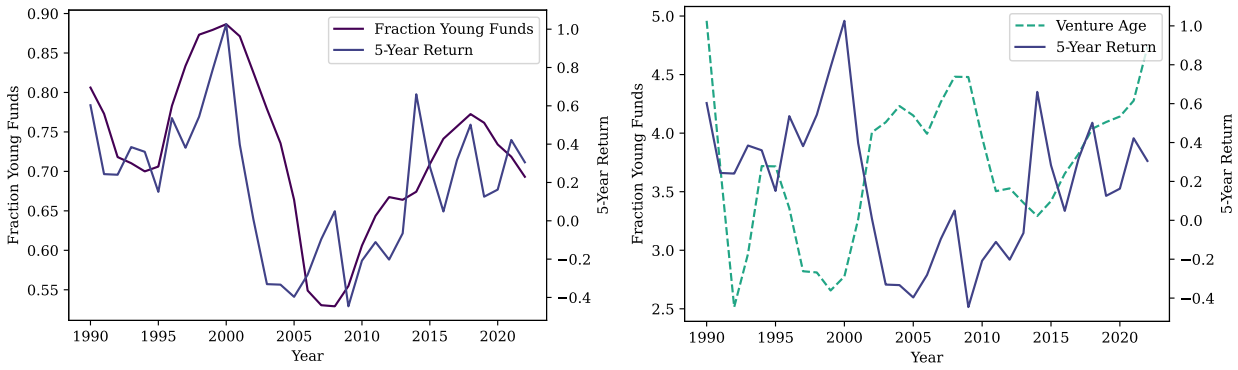

*Notes:* To make this figure, I calculate the number of first-time deals by a fund into ventures and calculate the cumulative fraction from years 0 to the last deal. Then, I calculate the mean and standard error across funds. Figure is discussed in Section 6.

**Figure A6:** Relationship Between Fund Age and Number of Previous Funding Rounds




*Notes:* On the y axis, I plot the average number of previous funding rounds that a venture has already raised when its raises a first-time investment from a fund. The x axis reports the fund age when the fund makes its first-time investment into the venture. Bars report the standard error across funds. The Figure is discussed in Section 6.

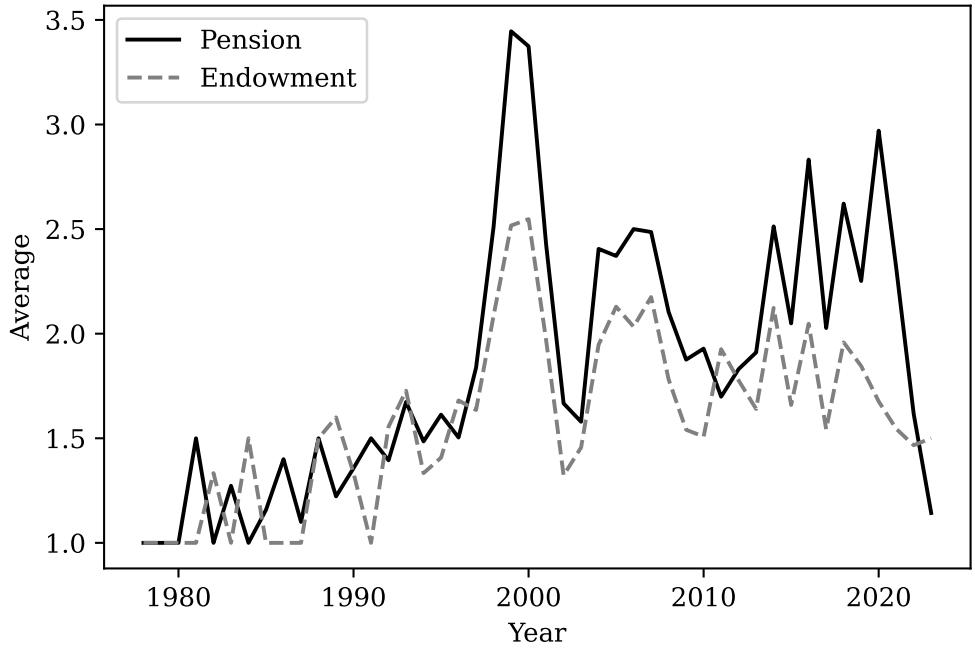
**Figure A7:** Instrumental Variable: 5-Year Return on the S&P 500




**(a)** First Stage **(b)** Reduced Form

*Notes:* Figure report graphs corresponding to the first stage and reduced form from the instrumental variable analyses presented in Section 6.1. The left panel shows the relationship between the endogeneous variable, fraction of young funds available in each year (in purple), and the 5-year S&P 500 return (in blue) that I use as the instrument. The right panel shows the reduced form relationship between the age of the venture at the time of the deal (dashed light blue) against the instrument.

**Figure A8:** Average Number of Funds Invested In by LP Institution Type




*Notes:* Figure reports the average number of funds that a Limited Partner (LP) invests in every year between 1980 and 2022 for two different institutional times of LPs: endowments and pension funds. The solid line presents the averages for pensions whereas the dashed line presents the same averages for endowments. Table A19 reports the counts of LPs of each type of institution.