Public R&D Spillovers and Productivity Growth

Arnaud Dyèvre
LSE

June 27th, 2024

[Latest version]

ABSTRACT. Does the source of Research and Development funding, public or private, matter for aggregate productivity growth? Using a novel firm-level dataset with patent and balance-sheet information covering 70 years (1950-2020), I estimate the impact of the decline in public R&D in the US on long-run productivity growth. I use two instrumental variable strategies–a historical shift-share IV and a patent examiner leniency instrument–to estimate the impact of the decline in public R&D on the productivity of firms through technology spillovers. I find that a 1% decline in public R&D spillovers causes a 0.03 to 0.08% decline in firm TFP growth. Public R&D spillovers appear to be two to three times as impactful as private R&D spillovers for firm productivity. Moreover, smaller firms experience larger productivity gains from public R&D spillovers. I calibrate a model of growth with heterogeneous firms which suggests that the decline in public R&D can explain around a third of the decline in TFP growth in the US from 1950 to 2017, and a third of the rise in size inequality between firms over the same period.

Key words: Growth, Firm heterogeneity, R&D, Productivity, Technology spillovers, Patents
JEL codes: O31, O32, O33, O38, D24, L25

Department of Economics, LSE, a.dyevre@lse.ac.uk, arnauddyevre.com
First version: October 15th, 2023.
For their invaluable, patient and supportive guidance, I am indebted to Arun Advani, Xavier Jaravel, Ben Moll and John Van Reenen. For insightful feedback and discussions, I thank participants in the Macroeconomics, Public Economics and EEE seminars at LSE, the Alan Turing Economics Data Science workshop, the Warwick PhD conference, the Concordi 2023 conference and WEFI. In particular, I extend my thanks to Ufuk Akcigit, Marco Bellifemine, Maarten De Ridder, Matthias Doepke, Gaia Dossi, Joe Hazell, Ethan Ilzetzki, Peter Lambert, Camille Landais, William Matcham, Filippo Mezzanotti, Javier Miranda, Jane Olmstead-Rumsey, Daniel Reck, Hugo Reichardt, Yannick Schindler, Oliver Seager, Kate Smith, Johannes Spinnewijn, Aditya Soenarjo, Abhijit Tagade, Alexander Whalley and Alwyn Young for their helpful comments.
This research benefited from financial support from STICERD (grants #107728 and #108954), which is gratefully acknowledged.

1. INTRODUCTION

The history of American technological progress is rich with examples of successful applications of government-funded research to the wider market economy. For instance, the US Department of Energy pioneered the development of lithium-iron batteries in the 1970s, and today's fast-growing vertical farming industry builds upon technologies first developed in the 1990s by NASA to grow plants in space. These public-to-private technology spillovers have been celebrated by advocates of a state-led approach to innovation. However, many see them as cherry-picked examples of an inefficient allocation of resources away from the private sector.

In spite of an extensive body of work on the topic of spillovers and growth, the impact of the decline in public R&D on productivity has remained an open question for three reasons. First, studying public-to-private spillovers at the firm level over 70 years is demanding in terms of data, and existing panels of firms matched to their innovations (usually measured by patents) are inadequate. These existing panels (i) are either too short, or (ii) do not contain sufficient information on who is funding R&D, or (iii) do not have measures of productivity at the firm level. Secondly, comparing the impact of public and private spillovers in a unified, causal econometric framework has not been attempted, perhaps because of the difficulty of finding plausible identification strategies for the impact of public R&D spillovers. Lastly, linking the impact of public R&D spillovers on firms to the aggregate consequences on the national economy requires a cautious treatment of general equilibrium effects.

In this paper, I address these challenges empirically and theoretically. I combine a newly assembled panel of firms matched to patents over seven decades (1950-2020) with two novel instrumental variable strategies to estimate the causal impact of public-to-private and private-to-private spillovers on firms' long-term outcomes. I then use the estimated spillover elasticities to calibrate a general equilibrium model of growth with heterogeneous firms. From these exercises, four key findings emerge.

The first key finding is that public R&D is different from private R&D, in particular in how much closer to science it is. I show that, even after controlling for differences in inputs into the research process, public R&D patents are more than twice as likely to rely on scientific publications than private R&D patents. Furthermore, I use a new measure of how 'ahead of its time' a patent is to show that public R&D patents are more likely to open new technological fields. These public R&D patents are also cited across a wider array of patent classes. Finally, they tend to be disproportionately cited by small firms. These facts suggest that publicly-funded patents embody ideas that are less appropriable by the original inventor and are therefore more likely to spill over to the rest of the economy.

The second key finding is that public-to-private spillovers have a large and positive causal impact on firms' productivity and innovative effort. Identification comes from a historical shift-share instrumental variable setting (SSIV), where I combine firm-level shares of exposure to R&D funded by US federal agencies with R&D funding shocks induced by geopolitical factors (such as wars, the Space race, the 1973 oil shock, etc.). Exposure shares are defined by the overlap in technologies in which a public agency and a company are active. The identifying assumption is that firm-level outcomes are orthogonal to the federal funding shocks conditional on time, industry, geography and lagged firm controls. As such, the identification relies on a quasi-experimental SSIV approach with exogenous *shocks* (Borusyak *et al.*, 2022).[1] I obtain historical estimates of the elasticity of impact of an increase in exposure to public R&D on long-term firm outcomes such as productivity, patent production, own R&D and sales over a long period (1945 to 2005). My estimates suggest that a 1% increase in exposure to public R&D causes a 0.025% rise in firm-level productivity. Additionally, public spillovers are more potent for smaller firms, perhaps because these firms have fewer resources to do in-house R&D (Acs *et al.*, 1994). As such, a decline in public R&D may be one of the causes of the rising inequality between firms and the growth of large firms.[2]

The third finding is that public R&D spillovers are between two to three times as impactful as private R&D spillovers for firm productivity. To compare the magnitude of public and private spillovers, I turn to a second identification strategy. I exploit the random allocation of patent applications to patent examiners of varying leniency to create measures of exposure to technology spillovers driven uniquely by this 'patent lottery'. This instrument is inspired by earlier work on judge leniency (Kling, 2006) and has been extensively used in the innovation literature (Gaule, 2018; Sampat and Williams, 2019; Feng and Jaravel, 2020). In contrast to previous studies, I use the patent lottery to instrument a firm's *exposure* to spillovers rather than its own patent grant decision. The identification assumption is that the variation in leniency at the examiner level is not correlated with the outcomes of firms that benefit from the spillovers of the reviewed patents. Previous evidence on the quasi-experimental assigment of applications to examiners suggest that this assumption is likely to hold (Lemley and Sampat, 2012), and I find support for it in the data. The advantage of the patent leniency instrument is that it allows me to estimate the causal impact of both public and private spillovers within the same econometric setting.

Finally, I find that the large decline in US public R&D matters quantitatively for aggregate TFP growth and inequality between firms. I build a general equilibrium, heterogeneous agent model

[1] I follow the latest literature in applied econometrics to implement this SSIV design (Adão *et al.*, 2019; Borusyak *et al.*, 2022) and use conservative, exposure-robust standard errors that take into account the correlation of firms' errors exposed to a similar set of federal agencies.

[2] Kwon *et al.* (2022) provide evidence that inequality between American firms, in sales and assets, has been increasing for most of the 20th century and, in particular, since the 1960s.

of growth in the spirit of Luttmer (2007) and Jones and Kim (2018) to quantify the macroeconomic implications of the decline of public R&D on firm productivity growth and the rise of superstar firms. In the model, R&D is performed by firms and by the government who levies taxes on firm profits to fund its R&D expenses. The model yields two key insights. The first is that aggregate productivity growth increases in the strength of spillovers while inequality between firms is decreasing in the strength of spillovers. The second insight is that there is a unique growth-maximizing corporate tax rate for growth. This tax rate is high enough to support the funding of public R&D but low enough to not discourage private innovation by firms. To go from my microeconomic evidence to general equilibrium conclusions, I use the elasticities obtained from my two empirical strategies to calibrate the model. The model suggest that the large decline in public R&D in the US in the second half of the 20$^{\text{th}}$ century may account for a third of the observed decline in aggregate TFP since the 1950s and a third of the rise in inequality of productivity between firms.

Related work. This paper relates to three strands of literature; the first of which is the voluminous set of applied papers on the importance of technology spillovers for innovation and productivity. Since the review of empirical studies by Griliches (1992) at least, it is recognized that spillovers from firms' R&D are common and economically significant. Estimates of the wedge between the private and social returns of corporate R&D suggest that social returns are two to four times as big as private ones (Bloom *et al.*, 2013; Lucking *et al.*, 2019). The literature has mostly focused on spillovers from firms' own R&D to other firms,[3] but recent work has shown that spillovers from the public funding of corporate R&D are also substantial. In two important contributions to this line of research, Azoulay *et al.* (2019) and Myers and Lanahan (2022) exploit quasi-experimental variation in federal agency funding rules to estimate the impact of public R&D grants on firms' own innovation and spillovers. Both studies conclude that spillovers from public R&D grants to firms are large: firms typically capture at most half of the returns of their own innovation.[4] This paper brings complementary evidence about the importance of public spillovers and extends this line of work in four main ways. First, I directly compare the impact of public and private spillovers within a unified econometric framework. Second, I go beyond specific agency programs and time periods by exploiting variation in spillovers across all patent-filing agencies and, for the

[3] Notable exceptions include Jaffe (1989), Belenzon and Schankerman (2013) and Bergeaud *et al.* (2022a) who study knowledge flows from academia to businesses, as well as Moser *et al.* (2014) and Iaria *et al.* (2018), who study spillovers within academia.

[4] Azoulay *et al.* (2019) find that a \$10 million increase in NIH funding generates 1.4 patent in the medical area targeted by the grant. But, importantly, it generates 2.2 additional patents in different areas (estimates from columns 4 and 5 of table 8, p. 145 in Azoulay *et al.*, 2019). Myers and Lanahan (2022) confirm this order of magnitude: firms capture only between 25 and 50% of the patent-based value of their publicly-funded R&D.

historical SSIV, variation from 1945 to 2010. Third, I use publicly-funded R&D in its broadest sense, regardless of who performs it. In other words, firms, universities and government labs are all included among the performers of publicly-funded R&D.

Moving from the micro-empirical evidence to the aggregate level, this paper also relates to the macro literature on idea-based growth, which has highlighted the central role of knowledge spillovers in driving aggregate growth (Romer, 1990; Jones, 1995; Jones and Williams, 1998; Lucas, 2009).[5] The central tenet of these models is that ideas are special inputs into a production function: they are non-rivalrous, and as such give rise to increasing returns (Jones, 2022). I show that while ideas generated by public or private R&D are both non-rival, they differ in how excludable they are: public R&D ideas are less excludable and therefore less appropriable. This lack of appropriability stems in large part from the fact that public R&D ideas are more fundamental. To my knowledge, this paper is the first to document this difference in appropriability between public and private R&D.[6] This point has important consequences for ideas-based growth models: public and private R&D need to be modelled separately because the spillovers they generate differ. I use my estimated elasticities to calibrate a model of aggregate growth with spillovers. In doing so, I provide a micro-to-macro framework that bridges the gap between the productivity literature on spillovers and macro models of growth. A contribution of this paper is to provide a tight theoretical link between idea-based models of growth and the econometric framework used by micro-empirical studies of firm growth. In addition, this work speaks to a few recent macroeconomics papers showing that reduced spillovers from market leaders to followers can worsen inequality between firms (Akcigit and Ates, 2019; Olmstead-Rumsey, 2022). My results suggest that reduced spillovers from public R&D to small firms are another potential explanation of the rise in firm inequality.

Finally, the present work contributes to the burgeoning literature about the role governments may play in driving productivity growth, either through demand shocks (Ilzetzki, 2022; Antolin-Diaz and Surico, 2022; Belenzon and Cioaca, 2022) or through large R&D expenditures (Kantor and Whalley, 2022; Fieldhouse and Mertens, 2023; Moretti *et al.*, 2023).[7] My work more directly relates to the second set of papers and complements them. While these papers focus on public R&D expenditures, I directly compare the potency of public and private spillovers for productivity growth. Moreover, I am leveraging detailed firm-level, balance-sheet data to test a wide

[5]See Buera and Lucas (2018) for a review of models of idea flow and growth. See Jones (2022) for a semi-endogenous growth perspective on the literature.

[6]See Akcigit *et al.* (2020) for a related point about basic versus applied R&D and Trajtenberg *et al.* (1997) for a comparison of university and corporate patents.

[7]In addition to academic papers, several general public books have collected case studies to make the case for a more central role for the government in pushing innovation forward. See for instance the books by Mazzucato (2015), Janeway (2018) and Gruber and Johnson (2019).

array of firm outcomes and uncover important treatment effect heterogeneity of public spillovers across the firm size distribution. Kantor and Whalley (2022) and Fieldhouse and Mertens (2023) conduct their analyses at the county and national levels, respectively.[8] Moretti *et al.* (2019) provide some firm-level evidence that businesses that receive government R&D increase their own R&D spending (and eventually experience higher productivity), but they do not investigate the role that technology spillovers play in this process.

The paper is structured as follows. In section 2, I briefly describe the novel dataset of publicly listed firms matched to patents that I use, before documenting stylized facts about patents funded by public R&D in section 3. Section 4 describes my two empirical IV strategies and their results are discussed in section 5. I present a model of growth through heterogenous firms and spillovers in section 6. The results of the calibration exercise are further discussed in that section. Section 7 concludes. Additional results, data description and proofs are relegated to the appendices.

2. Data

Studying technology spillovers at the firm level over 70 years is demanding in terms of data. Previous studies have been limited by panels of firms matched to patents that extend for at most 35 years.[9] This is inadequate to study the relevance of spillovers for growth from 1950 to 2020, the period during which public R&D has declined in the US. In this section, I describe the panel of publicly listed firms matched to patents that I assembled with a co-author (Dyèvre and Seager, forthcoming), and that I use in this paper. This panel spans seven decades and is the longest of its sort, doubling the time coverage of previous efforts (Arora *et al.*, 2021b). Importantly, it dynamically re-assigns patents to their current owners following corporate restructuring events (mergers, acquisitions, de-listings and spinoffs). The data is freely available to use for academic purposes and can be downloaded here: github.com/arnauddyevre/compustat-patents. A more detailed description of the data is available in Appendix B, and in Dyèvre and Seager (forthcoming).

Firm characteristics. Annual firm-level data come from Compustat North America, covering all firms publicly traded on a North American exchange. My final sample of firms consists of observations with employment, capital investment, operating income before depreciation and 4-digit SIC sectors. Using data on publicly listed firms has two advantages and one limitations. On the

[8]In spite of different methodologies and units of analyses, I obtain elasticities close to the ones reported by Fieldhouse and Mertens (2023): they report 0.2% increases in aggregate TFP, following a 1% increase in government R&D. They also find that the elasticity of *output* to government R&D is around 0.12, close to the elasticity of productivity (value added per worker) of 0.17 that I find at the firm level.

[9]Patent data alone cannot be used to study the impact of spillovers on firms because it lacks information on firm outcomes such as sales, employment and productivity. To my knowledge, the longest panels used to study spillovers are those created by Arora *et al.* (2021a) which runs from 1980 to 2015, Lucking *et al.* (2019) from the early 1980s to 2006 and Akcigit and Kerr (2018) from 1982 to 1997.

positives side, using Compustat data enables me to create a decades-long panel of firms. Secondly, Compustat has been extensively used in the innovation literature (Bloom *et al.*, 2013; Arora *et al.*, 2021b), which enables one to compares the results of the present paper to earlier work. A limitation of this data is that Compustat firms are not representative of the entire American economy. They are typically much larger than other businesses. The findings of this work, and in particular the results about firm heterogeneity, need to be taken with this caveat in mind. Nevertheless, conclusions drawn from this work can be informative about the wider economy due to the economic importance of Compustat firms in the aggregate economy. Estimates of their importance show that they account for 26% of US employment and 44% of its GDP (Dinlersoz *et al.*, 2018).

Patents. Patent information comes from the US Patent and Trademark Office (USPTO). For patents granted after 1975 and their citations, the data comes from Patentsview, the USPTO prime portal for patents granted over 1976-2022. A key feature of Patentsview is that assignees, locations and inventors' names are carefully disambiguated. For instance, patents assigned to 'IBM' and 'International Business Machines' are correctly assigned to the same firm. For patents granted before 1975 and their citations, I use the data scraped from the original patents files by Fleming *et al.* (2019), henceforth FGLMY. Lastly, I use historical CPC technology classes at the time of filing from Bergeaud *et al.* (2022b) and the USPC technology classes from PatentsView.

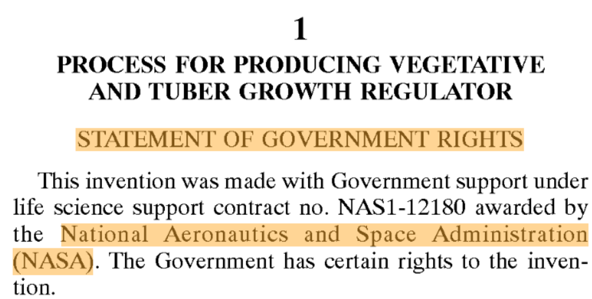
Patent data is an imperfect measure of innovation and appendix B elaborates on these limitations. However, it has been shown that patent counts correlate strongly with innovative inputs (R&D expenditures, number of inventors and scientists), other measures of innovative outputs (inventions rated by scientists) and proxies of firm performance (productivity, etc.). Moreover, while not all firms file patents, patents are a way to protect intellectual property that is extensively used by large firms (Mezzanotti and Simcoe, 2023) like the publicly listed firms in Compustat. Following the literature, I rely on patent data to quantify innovative outputs and on the overlap between patent technologies a to measure exposure to innovation.

Matching firms to patents. No unique firm identifier can serve as a joint between the balance-sheet data in Compustat and the USPTO patent data. Linking firms to patents must thus rely on matching company names to patent assignee names. Dyèvre and Seager (forthcoming) use a combination of string cleaning/homogenization, automated string matching, careful manual matching and reliance on the previous efforts of Arora *et al.* (2021b) to match firms to patents. They then rely on data from SDC Platinum, the Center for Research and Security Prices (CRSP), WRDS Company Subsidiary Data, historical data in Lev and Mandelker (1972) and manual searches to introduce dynamic reassignments of patents across firms, over time. Dynamic reassignment of patents is essential to obtain an accurate picture of firms' innovativeness at any point in time: patents indeed change hands over time through mergers, acquisitions and sales of subsidiaries.

The final matched dataset consists of 9,961 unique firm identifiers ('gvkeys') observed between 1950 and 2020 matched to 3.1 million unique patents. This is the most comprehensive dynamic dataset of Compustat firms matched to patents of its kind. Only a subset of these patents and firms are used in this paper because I need data on firms over at least 10 years to calculate my outcomes of interest and firms' exposures to spillovers. Appendix B and Dyèvre and Seager (forthcoming) provide more details about the matching procedure and compares the final dataset with existing alternatives such as Kogan *et al.* (2017) and Arora *et al.* (2021b).

Government-funded innovation. I define patents to be financially supported by the US government if they are assigned to a government entity ('direct assignee') or if the non-government assignee of the patent has received federal funding for the development of the innovation ('supported assignee'). Direct assignees are readily identified in PatentsView (post-1975) and FGLMY (pre-1975).

For supported assignees who are not government agencies, I use two data sources to identify government support. For patents filed after 1980, I rely on the 'government interest' variable created by PatentsView. The variable is derived from the text of patents whose assignees are required to disclose if they have received federal funding that contributed, even partially, to the innovation. An example of such disclosure is included in Figure 1, which shows an excerpt from a NASA-supported patent. This requirement comes from the Patent and Trademark Law Amendments Act of 1980–also known as, and henceforth, Bayh-Dole Act. It covers grants to firm, to universities and to NGOs, as well as procurement contracts between the government and any private or academic party. For patents granted before the Bayh-Dole Act, I use the government interest tag from Fleming *et al.* (2019). This tag comes from machine-read patent text where acknowledgement of government funding is reported.


1

PROCESS FOR PRODUCING VEGETATIVE AND TUBER GROWTH REGULATOR

STATEMENT OF GOVERNMENT RIGHTS

This invention was made with Government support under life science support contract no. NAS1-12180 awarded by the National Aeronautics and Space Administration (NASA). The Government has certain rights to the invention.

FIGURE 1. Example of a statement of government interest mentioning NASA – patent #5,992,090

Recent work by Gross and Sampat (2024) has shown that inferring government interest from the patent text or the Bayh-Dole disclosure statements, as I do above, can miss some relevant patents. In particular, 'license' patents which are funded by the government but assigned to non-government entities can be poorly covered, especially in the the 1950s and 1960s. I therefore complete the PatentsView and FGLMY datasets by Gross and Sampat (2024)'s government patent register.

Patent examiners' leniency scores. To create the examiner leniency instrument, I use data on all patent applications filed with the USPTO from 2001 to present days. The USPTO provides data on applications through its Patent Examination Research Dataset (PatEx), which includes information on special technology classes used for the allocation of applications to examiners called 'art units'. Crucially, this data contains the names of the patent examiners that I use to uniquely identify them.[10]

Department and Agency-specific funding. Historical data on R&D outlays by US agencies comes from the budget tables of the White House's Office for Management and Budget (OMB). This dataset needs to be completed because some departments that have historically funded R&D activities are not included in the White House R&D tables, like the Department for Veterans Affairs through its 'VA Technology Transfer Program' for instance.[11]

I fetch the additional R&D budgets of agencies not covered by the historical tables by cleaning a dataset of all government outlays available in the supplementary materials provided by the OMB, known as a the Public Budget Database. I isolate the R&D-specific outlays by performing a substring search among the 'Bureau Name' and 'Account Name' fields; I look for variations of substrings such as 'INNOVATION', 'RESEARCH' and 'TECHNOLOGY'. When data on R&D funding is available both in the series provided by the historical tables and in the detailed outlays, there is a very good overlap between the two series, as can be seen in panels A.7-A.9 in the Appendix. When both series are available, the series from the historical is used. Finally, I manually collect R&D data for the Department of Veterans Affairs and the Small Business Administration from Congressional Research Service reports. Values are deflated and expressed in 2020 dollars.

3. Stylized facts on public R&D patents

In this section, I use all 8.2 million patents granted from 1976 to 2020 by the USPTO to document three key characteristics of public R&D patents: (1) they rely more on science, (2) the knowledge they encode tends to be more ahead of its time and (3) they generate more spillovers, especially to smaller firms.[12] These differences with privately-funded patents have important consequences on the frequency and strength of spillovers. While a complete investigation into the causes of these

[10]The data is freely available on the USPTO website (www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair). Miller (2020) provides a comprehensive overview of the data.

[11]The Department of Veterans Affairs is active in financing and commercializing technologies that can benefit Veterans' wellbeing. Most of the patents financed by the Department of Veterans Affairs are medical patents and are typically jointly filed with inventors in academia (Department of Veterans Affairs, 2022).

[12]The controls I use in my specifications come from data only available in the post-1975 tranche of patent data. I therefore discard the 1950-1975 patent data for the analysis of this section.

differences is beyond the scope of this paper, I briefly discuss plausible reasons at the end of the section.

To test for differences between public R&D and private R&D patents, I regress some outcomes of interest y_i at the patent level, on an indicator variable equal to 1 if a patent is publicly-funded *i.e.* assigned to a 'direct assignees' or a 'supported assignee', and a comprehensive array of controls $\mathbf{X}_i$. publicly-funded patents can be the result of R&D performed in government labs, in universities, in firms or any combination thereof provided that at least part of the R&D money came from public sources. Formally, in the figures below I report the $\hat{\beta}$ coefficients and their 95% confidence intervals from the following regression, for a gradually more comprehensive set of controls $\mathbf{X}_i$:

$$y_i = \alpha + \beta \times \mathbb{1}[\text{patent } i \text{ is publicly-funded}] + \mathbf{X}_i\gamma + \varepsilon_i \tag{1}$$

Evidently, the $\hat{\beta}$ coefficients cannot be interpreted as causal. This exercise is however informative about the differences between public and private R&D, as seen through the lens of patented innovations. Heterogeneity results across years, performer and funders of public R&D are presented in Appendix C.2, along with robustness checks using alternative dependent variables.

3.1. Fact 1 - Public R&D patents are more reliant on science.

The most important difference between publicly-funded patents and privately-funded ones is in how much more reliant on science public patents are. To measure a patent's reliance on science, I follow the common practice in the innovation literature to use patent citations to proxy for knowledge spillovers.[13] Reliance on science is defined here as the share of a patent's backward citations directed to the scientific literature. Previous empirical work has shown that citations to the scientific literature are correlated with actual reliance on science in industrial R&D. For example, using the Carnegie Mellon Survey of the Nature and Determinants of Industrial R&D, Roach and Cohen (2013) document that there is a strong correlation at the industry level between the share of patent citations directed to scientific publications and the extent to which research lab managers report relying on science.

To calculate the share of citations to science, I rely on data compiled by Marx and Fuegi (2022) on non-patent citations. Using specification (1), I find that public R&D patents tend to rely more on science than private patents. The results are shown in Figure 2a, where I report point estimates and 95% confidence intervals for the β coefficients across a suite of specifications with successively

[13]Patent citations can be a noisy proxy for knowledge spillovers. But they have been shown to be strongly associated with actual spillovers, as reported in surveys by the inventors themselves. Jaffe *et al.* (2000), for instance, use a survey of inventors to show that patent citations often capture direct communications between inventors, word-of-mouth and the simple act of reading the cited patent. Moreover, citation patterns also correlate strongly with the movements of scientists between assignees citing each other's patents in my data. This suggests that one of the key channel through which the exchange of ideas operate–the mobility of inventors–is captured to some extent by citation flows. See section B for a discussion about the merits and drawbacks of relying on patent citations to measure spillovers.

more exhaustive controls. In my fullest specification, I control for 700 CPC patent class dummies, the productivity of inventors, the productivity of the entity who owns the patent and the estimated total wage bill of inventors. Standard errors are clustered by year of application and by patent class. I find that only 6% of citations made by private R&D patents are directed toward scientific papers. In contrast 22% of citations made by public R&D patents are (+267%). Appendix C.2 shows that this difference is stable over time and it persists even within R&D performers *i.e.* firms' and universities' innovations are more reliant on science when their funding is public than when their funding is private.

One interpretation of this greater reliance on science is that publicly-funded innovations tend to use knowledge that is more basic or more fundamental. Basic research is defined by the OECD 'Frascati manual' as 'experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view' (2015, p. 45). This definition is used by many public science agencies in their R&D surveys, including the US National Science Foundation. While there are both basic and applied pieces of scientific work, it is reasonable to assume that science articles tend to be more detached from practical applications and commercialization of ideas than patents, whose purpose is indeed to protect the profits of an invention. By relying on more fundamental knowledge, publicly-funded patents may themselves embody more fundamental knowledge. Two pieces of evidence support this interpretation. First, in appendix C.2, I also show that the number of independent claims made by publicly-funded patents is greater, on average. Patent claims delineate the scope of an innovation and establish which property rights the assignee is entitled to (Matcham and Schankerman, 2023). The larger this number, the least specific an innovation is. The number of independent claims can therefore be seen as a measure of the generality of a patent. Because basic innovations have applications across many fields, a patent's generality can be seen as a manifestation of its basicness. Second, the breakdown of public R&D across basic research, applied research and development is very different from that of private R&D. Out of each dollar invested in public R&D by the American government in 2020, 33 cents were dedicated to basic research and 36 were dedicated to development. The remaining 31 cents were used to fund applied research. In contrast, a dollar of private R&D in 2020 funded mostly development (78 cents) and very little basic research (7 cents). This split is shown in the figures of panel A.10 in the Appendix. I observe the consequences of this divergence of focus in the patent data.

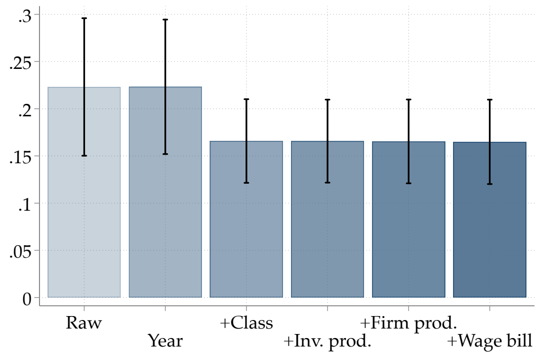
3.2. **Fact 2 - Public R&D patents are more impactful.** Secondly, to assess a patent's technological importance, I introduce a new metric of impact. I measure a patent's technological novelty by the number of years that separates its year of application from the date when it is reclassified into a newer patent class. Disruptive innovation, by definition, is hard to classify using existing taxonomies: patents that are re-classified into a newer, more relevant patent class after its introduction can therefore be thought as encoding knowledge that was 'ahead of its time'. I study the dynamic reassignment of patents to classes using the evolving US Patent Classification System (USPCS). It consisted of more than 450 classes and was in use from the early 19th century until 2013.[14] The USPTO needs to keep an up-to-date classification of technologies in order to assess the claimed novelty of patent application against existing prior art. Because of its important legal role, the USPTO had strong incentives to keep this classification relevant to the technological landscape of the time. After the introduction of a new patent class, all previously filed patents that are better described by the new class are *ex post* re-classified into the more relevant class. For instance, a patent filed in 1996 and protecting a technology that is relevant for the development of self-driving cars would be re-classified from, say, "Data processing: Vehicles, Navigation, and Relative location" (class 701) to "Data processing: Artificial Intelligence" (class 708) in 1998, when the latter is created. This patent would have contributed to open a new technological field two years before this field is recognized by the USPTO. The list of USPC classes thus offers an interesting vantage point into the development of new knowledge. Figure 13 in the appendix shows the cumulative count of USPC patent classes over time and indicates when some selected technologies are introduced.[15]

As shown in Figure 2b, I find that publicly-funded patents tend to be 6% more likely to be 'ahead of their time' than privately-funded patents (baseline probability with full controls: 0.31), even after controlling for the R&D effort, as proxied by the wage bill of innovators, that goes into the creation of the patent. This suggests that publicly-funded patents are not more impactful because they are the result of more expensive research. Looking at the intensive margin, I restrict the sample to patents that are ahead of their time and compute the difference in average years between the typical public R&D patent and the typical private R&D patent. I find that public R&D patents are typically 1.25 more years ahead than private patents (+19%). This result is reported in the Appendix. When using other common measures of impact such as forward citations and the Kelly *et al.* (2021) metric of breakthrough patents, the results also suggest that publicly-funded patents are more impactful, even after controlling for R&D effort (see Appendix C.2).

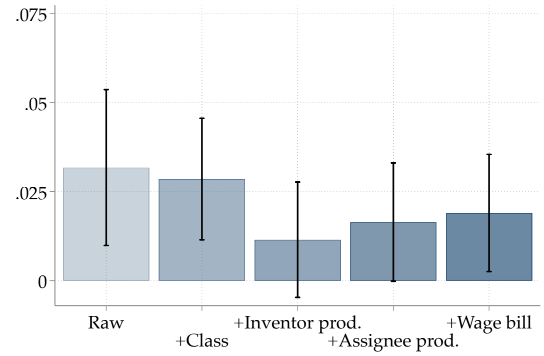
3.3. **Fact 3 – Public R&D patents generate more spillovers.** The last fact I document pertains to the breadth of spillovers from public R&D. I find that public R&D patents tend to generate spillovers across a wider range of patent classes. The excess number of classes across which a

[14]The Cooperative Patent Classification (CPC) system, jointly developed by the USPTO and the European Patent Office, replaced the USPC in 2013. While the CPC is also regularly updated, its late introduction makes it less interesting to study patent re-classification over the long term.

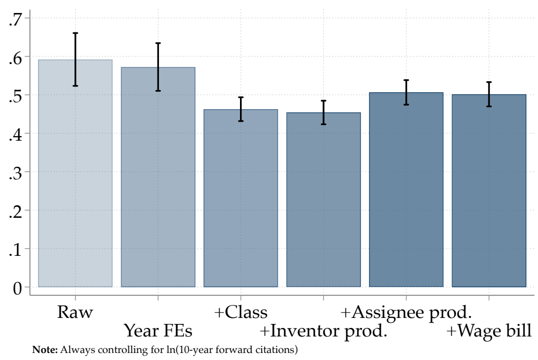
[15]Raw data stored at the following link arnauddyevre.com/files/USPC_classes_years_established.pdf. Csv file available at arnauddyevre.com/files/timeline_detail_classes.csv



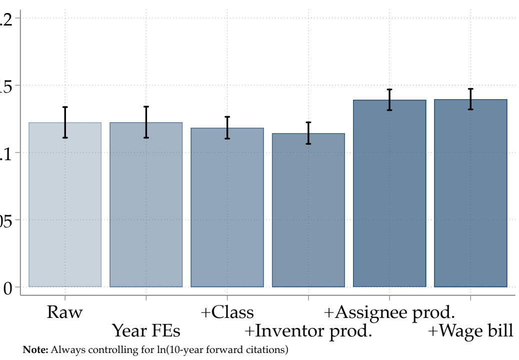
(A) Fact 1 – Share of backward citations to scientific papers



(B) Fact 2 – Patent is 'ahead of time'



(C) Fact 3.1. – Number of classes forward-citing the patent



(D) Fact 3.2 – Share of small firms citing the patent

FIGURE 2. Stylized facts about public R&D patents

Notes: The figures show the β coefficients and their 95% confidence intervals from specification 1, where the dependent variable is the number of years separating a patent from the creation of the USPC class it is eventually assigned to (A), the share of citations made by the focal patent to scientific literature (B), the number of CPC patent classes citing the focal patent and the share of small firms among the assignees citing the focal patent (D). The construction of the variables 'inventor productivity', 'assignee productivity' and 'wage bill' is described in Appendix C. The sample sizes are $N_A = 8.2m$, $N_B = 8.2m$, $N_C = 5.2m$ and $N_D = 5.2m$. In the 'ahead of time' regressions, I am not controlling for years and patent class jointly: the overlap between historical USPC classes and CPC classes used as controls is high and controlling for CPC classes and year leaves very little variation in y_i.

public R&D patent is cited is displayed in figure 2c. After controlling for many observables, public R&D patents tend to be cited by 0.5 more classes, from a baseline of 2.38 for the average private patent (+22%).[16] To disentangle the effect of the breadth of a patent from that of its technological impact, I also control for the log number of total citations received by the focal patent. The wide applicability of the knowledge encoded by public R&D patents is likely to stem from them being more fundamental, as documented in fact 1. This finding has important implications for the

[16]This finding echoes that of Babina *et al.* (2023) who find that patents funded by federal grants are more 'general'. Generality is defined as $1 - \sum_j c_{ij}^2$ where c_{ij} is the share of citations to patent i coming from class j.

appropriability of public research, which appears more limited than that of private research, and will be a key driver of the dynamics of the model.

Moreover, public R&D patents generate spillovers to a different distribution of firms than private R&D patents. In panel 2d, I report estimates from regression (1) where y_i is the share of citations received by patent i from 'small' firms, defined as firms with fewer than 500 employees. The data on firm size comes from patent applications, where firms are asked to report their size in order to determine the patent renewal fees they need to pay. Smaller firms face lower fees. Patents funded by public R&D money appear to be more likely to be cited by smaller firms: after controlling for the full suite of controls, I find that the share of small-firm citations to public R&D patents is 14 percentage points higher than for private R&D patents (+62%) suggesting that their technology spillovers are comparatively more relevant for smaller firms. This evidence is consistent with summary statistics reported by Azoulay *et al.* (2019), who find that small assignees (*i.e.* with fewer than 500 employees) are more likely to cite patents linked to NIH-funded research.[17]

One plausible interpretation of this finding is that smaller firms lack the resources and the incentives to perform basic research, unlike large companies such as DuPont, General Electric, IBM, Xerox or AT&T through Bell Labs which are prominent examples of firms with once dynamic basic research labs. Another interpretation is that university spinoffs through which academic researchers can develop commercial applications of their research have become more common, in particular after the passing of the 1980 Bayh-Dole Act that facilitated university patenting and licensing. Academic startups, because of their more agile way of doing business and close ties to university research, may have a comparative advantage in generating inventions, while established firms are better at exploiting innovations through development and commercialization (Arora *et al.*, 2018).

3.4. **Summary and discussion.** In summary, R&D funded by public money tends to be more of a *public good*: it is more impactful (as measured by citations, its ability to open new fields), more fundamental and less appropriable. These differences hold irrespective of who is *performing* the R&D, whether it is a university or a firm.[18]

Why is publicly-funded R&D different? Both the actions of the funder of public R&D (*i.e.* the government) and those of researchers receiving public funding offer explanations. Firstly, public R&D money tends to be much more heavily invested into 'basic' research, as can be seen in the

[17] Table 2, p. 133.

[18] Importantly, the stylized facts highlighted here are not a comparison of university and government lab patents versus corporate patents. Previous research like Trajtenberg *et al.* (1997) has for instance highlighted the relevance of the distinction between corporate and academic patents in determining the basicness and appropriability of patented technologies. In contrast, the results presented in this section and in Appendix C.2 reveal that the *source of R&D funds*, even within a university or a firm or a government matters for the impact, generality and appropriability of innovation.

histograms of panel A.10 in the Appendix.[19] This difference in the type of research being funded has consequences on the types questions being investigated, and eventually on the type of innovations being patented. Secondly, the incentives of researchers doing publicly-funded R&D may differ. Inventors doing publicly-funded research may be driven by prizes, publication-based promotion procedures and the satisfaction of having one's ideas widely used. See for instance the review by Williams (2012) on the effect of prizes in inducing innovation, Reschke *et al.* (2018) or Jin *et al.* (2021) for causal assessments of the importance of prizes in steering scientific research and Brunt *et al.* (2012) for their effect in industrial innovation. This interpretation echoes the findings of Babina *et al.* (2023), who use administrative data on university researchers matched to the funding composition of their grants (public or private) and find that researchers alter the trajectory of their research when their funding gets dominated by private funds. Their research becomes less open, less basic, more appropriable by the funder and of lesser academic quality.[20]

Is it due to selection? One might worry that the selection process of public R&D innovations that make it into patents is different than for private R&D. Researchers doing public R&D may be more conservative when deciding if the fruit of their research is worth patenting: they may be less interested in the money they can get from filing a patent for instance. As a result, the low impact, high appropriability and low basicness of private patents may simply be driven by a large volume of 'junk' corporate patents that do not exist in universities and government labs' patent portfolios. While this hypothesis is inherently hard to test, some evidence suggest that this may not be the case. Firstly, the conversion rate of patent applications into granted patents are similar for patents funded by private R&D and those funded by public R&D. Public applications are only 3 percentage points more likely to be converted than private applications (baseline: 83%). Secondly, when looking a citation-weighted patents, one diminishes the risk that the average quality of private patents is dragged down by low-quality patents. Only blockbuster patents, which are arguably very likely to clear the quality threshold for grant, matter in this exercise. When running the same analysis weighting patents by citations, the conclusions remain the same (results not reported). Also, looking at the distribution of patent citations, one finds an almost identical distribution for the bottom 90% of public and private patents. Thirdly, one may argue that 'junk' patents also

[19]The National Science Foundation and the OECD defines basic research as 'experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view' (Frascati manual, 2015, p. 45), while applied research is defined as 'original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific, practical aim or objective.'

[20]Relatedly, some studies have shown that necessity in periods of crisis can be a powerful catalyst of innovation by directing effort toward a common mission (Mazzucato, 2021): Ilzetzki (2022) studies the ramping up of US military aircraft production during WWII, Agarwal and Gaule (2022) looks at the redirection of clinical trials during the Covid-19 pandemic and Hassler *et al.* (2021) document the technical change that was spurred by the 1973 oil shock.

exist in the public R&D portfolio.[21] Finally, the regressions above are controlling for the effort put into each patent by including proxies of inventor's productivities, assignee productivities and the total wage bill of inventors on the patent. This creates comparisons between patents which have benefited from the same amount of research. Overall, there is very limited evidence that the differences between public and private R&D patents documented in this section are driven by different selection processes of innovations into patents.

4. Research designs

The previous section has shown that privately-funded R&D is different from publicly-funded R&D. This section lays out the econometric approaches I use to investigate the consequences of these differences for spillovers, firm growth and innovativeness. I first ground my estimating equation in the theory of knowledge production functions commonly used in empirical studies of spillovers (Griliches, 1979; Acs *et al.*, 1994), before discussing endogeneity issues. I then describe the two quasi-experimental IV strategies I use to estimate the causal impact of spillovers from government-funded research and privately-funded research.

4.1. From theory to data. To motivate the equation I am estimating, it is helpful to think of firms as being endowed with the following productivity process, which is at the heart of the model presented in section 6:

$$\dot{Z}_{it} = {E_{it}}^{\phi}\Gamma_{it} \quad \text{with} \qquad \Gamma_{it} := \left(\prod_a P_{at}^{s_{iat}}\right)^{\gamma} \left(\prod_f {P_{ft}}^{s_{ift}}\right)^{\varepsilon} \tag{2}$$

where E_{it} is the (flow) R&D effort of firm i at time t, ϕ is the elasticity of productivity growth ($\dot{Z}_{it}$) to R&D expenditures and Γ_{it} captures the spillovers to which the firm is exposed. Departing from previous research, I define Γ_{it} as being a composite term capturing spillovers from publicly-funded and privately-funded R&D that i benefits from. It is made of two Cobb-Douglas aggregators, one for each type of spillover: public spillovers come from agencies indexed by a and private spillovers come from firms indexed by f. P_{at} and P_{ft} are the (flow) patents of agency a and firm f, respectively. For each firm i exposed to patents funded by agencies, I remove from P_{at} the patents that are funded by a but filed by the focal firm i, if there are any.[22] In other words, focal firms are

[21]Some agencies like NASA have an explicit mandate to facilitate the translation of NASA's research into civilian development (through its Transfer Technology program and yearly Spinoff publication). While some of its patented innovations have had successful applications in civilian domains (such as NASA's research into LED light), others are simply using the patent system as a way to make these innovations known to the public and/or facilitate spillovers. See for instance the lunar module landing pad patent (#3,175,789) or this quite imaginative 'space spider crane' (#4,738,583)

[22]P_{at} is therefore a slight abuse of notation as it should also be indexed by i.

not exposed to their own innovation in my setting.[23] Correspondingly, i is not included in the set of spillover-generating firms indexed by f, although it may generate spillovers to other firms.

The shares s_{iat} capture the importance of agency a's knowledge production in firm i's spillover aggregator. They sum up to 1 within each type of spillovers and can therefore be interpreted as follows: $s_{i,\text{NASA}} = .25$ means that variation in NASA's knowledge mediates 25% of the variation in firm i's exposure to publicly-funded spillover and $\gamma \times .25$ of the variation in its productivity growth. Shares of exposure to privately-funded R&D, s_{ift}, are defined analogously as the importance of firm f in firm i's private spillovers. Importantly for my purpose, and in contrast with previous work, I allow the elasticity of productivity to exposure to public R&D, γ, to be different from that of private R&D, ε.

Taking logs, one can estimate equation (2) as:

$$\Delta z_{it} = \phi \underbrace{e_{it}}_{\substack{\text{own ln} \\ \text{R\&D flow}}} + \gamma \underbrace{\sum_a s_{iat} p_{at}}_{\substack{\text{exposure to} \\ \text{public R\&D patents}}} + \varepsilon \underbrace{\sum_f s_{ift} p_{ft}}_{\substack{\text{exposure to} \\ \text{private R\&D patents}}} + \epsilon_{it} \tag{3}$$

where $\Delta x_t := \ln(X_t) - \ln(X_{t-1})$. In what follows, I discuss the construction of the exposure variables. I also discuss the timing of measurement of the various empirical elements of equation (3). I have economized on notation here by indexing all variables by $t-1$ and t, but the timing of spillovers relative to their impact on productivity growth is important and is discussed later.

Shares of exposure. In line with previous work in the spillover literature, I calculate the shares of exposure s_{iat} following the methodology pioneered by Jaffe (1986) and subsequently used by Bloom *et al.* (2013) and Bloom *et al.* (2020). The Jaffe proximity metric relies on the overlap in technologies between two patent assignees to situate them in technology space. The more similar the distributions of patents of two assignees across technologies are, the closer these assignees will be according to the Jaffe metric and the more likely they will be to benefit from spillovers emanating from each other's innovations. Formally, I define $\mathbf{P}_i := (P_{i1}, P_{i2}, \ldots, P_{iN})$ as the $(1 \times N)$ row vector of shares of patents of firm i across the N technology classes in a given period. Time subscripts are omitted for readability. For instance, if a firm i holds only two patents, one in the 'Soilless cultivation' class (4-digit CPC class: A01G) and one in 'Devices for administering medicine orally' (A61J), then its technology signature vector will have 0 entries everywhere except for $P_{i,\text{A01G}} = P_{i,\text{A61J}} = .5$. $\mathbf{P}_a$ is defined analogously for agency a. The proximity between i and a is defined as the uncentered correlation between i and a's technology shares of patents:

[23] The R&D term in equation (2) already captures a firm's past innovative effort.

$$\widetilde{s_{ia}} := \frac{\mathbf{P}_i \mathbf{P}_a'}{\sqrt{\mathbf{P}_i \mathbf{P}_i'} \sqrt{\mathbf{P}_a \mathbf{P}_a'}} \qquad \in [0,1] \tag{4}$$

$\widetilde{s_{ia}}$ ranges from 0 (no overlap in technology signature between i and a) to 1 (identical shares of patents across classes). I calculate these exposure weights using patents over a period of 5 years, starting 5 years before firms' outcomes are observed. Therefore, to estimate the impact of spillovers on a firm's sales growth from t to $t+5$, exposure weights are calculated using patent data from $t-5$ to t. To define the share of exposure to a particular agency, I normalized the proximity metrics $\widetilde{s_{ia}}$ such that they sum up to 100% across agencies *i.e.* $s_{ia} := \frac{\widetilde{s_{ia}}}{\sum_{a'} \widetilde{s_{ia'}}}$. These shares of exposure are interacted with the log of patent production by agency a, p_{at}, to create the change in exposure to public spillovers. I define p_{ft} and s_{ft} analogously, as the patent production by firm f at time t, and the shares of exposure to firms indexed by f, respectively. I show in Figure 12 in Appendix B.5 that shares of exposure are very stable over time: the correlation in shares of exposure to public agencies measured over one five-year interval with shares in the next five-year interval is very high for the majority of shares, which are between 0 and .2.

An alternative to using technological overlap between entities to define shares is to instead rely on patent citations. This approach however has several drawbacks. The first is that patent citations are sparse; they only represent a tiny sliver of the knowledge base used in the creation of an innovation. The second is that patent citations can be a noisy signal of knowledge flows. Third, there are some solid microfoundations behind the use of the technological overlap as a measure of knowledge flow (see Bloom *et al.* 2013). Lastly, this makes my approach comparable to the literature.

Timing. Importantly, the timing of the dependent and independent variables in specification (3) needs to be informed by empirical evidence about the delays taken by spillovers to materialize. In particular, one must take a stand on the time it takes for an idea generated by an upstream knowledge producer (either a private firm or a public agency) to be converted into profitable product and services by downstream firms. This dynamic aspect of spillovers is, surprisingly, rarely discussed in microeconomic studies of spillovers. The evidence on the so-called 'invention-innovation' lags comes from a small literature that has used surveys, case studies, as well as bibliometric data on patents and academic papers. Its findings suggest that lags of around five years between the dissemination of an idea–*e.g.* through a patent or paper publication–and the introduction of a product or service that builds on it are common, with significant heterogeneity across industries.

Mansfield (1991) for instance surveyed R&D executives in American manufacturing firms who used extramural research findings in the development of their products or processes. The mean reported lag between the publication of a finding and the first commercial introduction of a product using it was 6.4 years. There is some heterogeneity across industries though: pharmaceutical firms experience the longest lags (10.3 on average), firms in 'Instruments' experience the shortest (4.2). Similarly, the National Science Board in the US reports that the mean time between the first conception of an innovation and the innovation itself is 7.2 years, for a sample of 500 academic innovations used in product or processes by American firms between 1953 and 1973 (National Science Board, 1975).[24] Mowery *et al.* (2015) present several case studies of academic innovations that have been successfully commercialized and offer a detailed description of their patent-to-product timelines. The co-transformation process, an important application of modern genetics, took between four and seven years to be used in biomedical firms' productions. The commercial development of LED lights using Gallium nitride–a semiconductor emitting light over a wide spectrum of colors–took between two and seven years. The glaucoma drug Xalatan took between nine and 14 years.[25] Another piece of evidence comes from Ahmadpoor and Jones (2017) who use the shortest lag between the publication of a paper and the publication of a patent that cites it as a measure of spillover delay. They find an average delay of 6.7 years.

Taken together, the findings of this literature suggest that, in spite of the heterogeneity in lags, spillovers from inventions to commercialization typically take between five and 10 years. Using patent production of the spillover-generating entities at t, and differences in the outcomes of interest of firms from t to $t+5$ (or flow patent production at $t+5$) thus appears warranted. This timing allows firms in my sample to be exposed to spillovers and to be impacted by them within a reasonable timeline so that I can observe changes in productivity. My own empirical work, presented later in the paper (Figure 4b), provides a justification for the lag between R&D investments by agencies and patent creation. The timeline shown in Figure 3 summarizes the timing used in the variable creation.

4.2. **Endogeneity.** If a researcher could run the ideal experiment to estimate (3), she would choose, at random, how many patents p_{at} and p_{ft} upstream agencies and firms are generating in year t. In such hypothetical case, the exposures to spillovers $\sum_a s_{ia} p_{at}$ and $\sum_f s_{if} p_{ft}$ would be orthogonal to the error ϵ_{it} by design. With this ideal experiment, the OLS regression of firm i's log productivity

[24] The report studies 500 'major' technological innovations defined as 'new products or processes embodying a significant technological change'. They include technologies like nuclear reactors, lasers and oral contraceptives. Interestingly, these lags tend to vary by country: the average is 3.6 years in Japan, 5.6 in west Germany, 6.3 in the UK and 7.4 in France (table 1-13 and figure 1-13 in the NSF report).

[25] These are all examples of lags between the dissemination of an innovation and its application by a firm, these are not lags between the production of science and productivity externalities accruing to firms relying on science. These science-to-firm lags are typically found to be much longer than innovation-to-firm. Adams (1990) estimate this lag to be of the order of 20 years, and Marx and Fuegi (2020) find that the average time lag between a patent application year and the publication year of the papers it cites is 17 years.

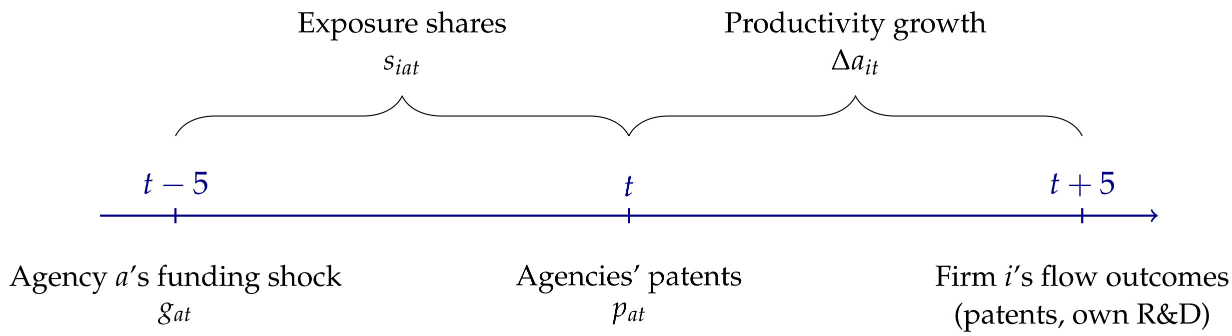


FIGURE 3. Timeline

Notes: The figure describes the timeline used to construct the data that I rely on for the estimation of (3). It is informed by the literature on the 'invention-innovation' lag reviewed in the main text of the paper. It also relies on the empirical exploration of the lag between funding shocks and patent creation shown in Figure 4.

change at time t on its exposure to federal and private innovation yields unbiased estimates $\hat{\gamma}$ and $\hat{\varepsilon}$.

Departing from the ideal experiment, firms' exposures to government-funded innovations may not be random and the exclusion restriction $\mathbb{E}[\boldsymbol{\epsilon}' \sum \boldsymbol{sp}|\boldsymbol{e}] = 0$ may not hold. The most likely threat to identification comes from correlated shocks to technologies that affect both the propensity of upstream agencies to innovate and the outcomes of downstream firms. Technological advances like the creation of the personal computer or the development of mRNA vaccines may present new R&D opportunities for the Department of Defense and the Department of Health and Human Services, respectively, while at the same time offering growth opportunities to IT and pharmaceutical firms exposed to these agencies. This type of correlated shock would bias OLS estimates upward and is a standard manifestation of the 'reflection problem' (Manski, 1993). Another manifestation of correlated shock would be government demand shocks that may increase R&D spending of an agency (like the DoD in period of war) and at the same time increase demand for firms who are both exposed to spillovers and government contractors (like defense firms).

In addition to correlated shocks, a second threat to identification comes from reverse causality. The government may be increasing some agencies' R&D because the productivity of a given sector has been disappointing. This could be the case of the health sector, which is exposed to research conducted by the various institutes of the Department of Health and Human Services, and whose productivity growth, by some accounts, has been lower than in the wider US economy (Spitalnic *et al.*, 2016).

Several choices are likely to limit the extent of these endogeneity concerns. Firstly, the choices of time periods used for the variable construction helps in alleviating both correlated shock and reverse causality issues. Technology spillovers are operating over relatively long time periods (between five and 10 years according to the literature reviewed in 4.1), while government demand shocks such as those caused by wars or pandemics are typically short lived and have immediate impacts on government contractors' performance. Antolin-Diaz and Surico (2022) find that impulse responses of government spending following military news are indistinguishable from 0 (at the 68% level) after five years.[26] In a careful causal analysis of a government demand shock on plants' productivity, Ilzetzki (2022) shows that demand-induced productivity increases in aircraft manufacturing plants starts decreasing 15 months after the initial shock with output per worker growth undistinguishable from 0 after 18 months (95% level).[27] Government demand shocks and government-generated spillovers are working on non-overlapping timeline: while the short run effects of an increase in government spending are due to demand, they are due to spillovers at longer horizon. Reverse causality issues are also unlikely to be serious because of the way in which standard policymaking is conducted: changes in agencies budget are most likely to be informed by *past* economic outcomes than economic outcomes in the future.

Secondly, to mitigate the impact of government demand shocks, I remove from my sample firms in sectors most likely to be exposed to these shocks. These sectors are: 'Guided Missiles & Space Vehicles & Parts' (SIC4 code: 3760), 'Aircraft' (3721), 'Search, Detection, Navigation, Guidance, Aeronautical Systems' (3812), 'Pharmaceutical Preparations' (2834), 'Wholesale-Drugs, Proprietaries & Druggists' Sundries' (5122), 'Services-Computer Integrated Systems Design' (7373), 'Ship & Boat Building & Repairing' (3730) and 'Biological Products' (2836). Their exclusion removes arms and aircraft manufacturers such as Lockheed Martin or Raytheon and all big pharmaceutical firms such as GSK and Pfizer.

Thirdly, one way to evaluate the extent of correlated shocks and reverse causality is to exploit the panel nature of my SSIV setting and conducting falsification tests using lagged outcomes. If the productivity growth of firms more exposed to spillovers is higher in the pre-period, this would be indicative of a violation of the exclusion restriction. I test for pre-trends and pre-levels in section 5 and find no evidence that more treated firms are different or on a different growth trajectory than less treated firms.

Lastly, to deal with unobserved heterogeneity, I assume that the error ϵ_{it} is the sum of a 2-digit-sector-specific fixed effect $\eta_{s(i)}$, a 5-year period fixed effect τ_t, a geography (=state) fixed effect $\lambda_{g(i)}$, and an idiosyncratic component (v_{it}) that I allow to be correlated across firms exposed to a similar set of agencies (Adão *et al.*, 2019) and heteroskedastic. In my fullest specifications, I also control for four lagged firm observables, in the matrix $\mathbf{X}_i$: capital stock, sales, employment and patent count, all in logs. The full structural equation of my SSIV setting is thus:

[26] Figure 1, first panel.

[27] Figure 8(b).

$$\Delta z_{it} = \phi e_{it} + \gamma \sum_a s_{iat} p_{at} + \varepsilon \sum_f s_{ift} p_{ft} + \eta_{s(i)} + \tau_t + \lambda_{g(i)} + \mathbf{X}_{it}\boldsymbol{\beta} + v_{it} \tag{5}$$

Because the model is in long differences, any firm-specific constant fixed effect will be differenced out. Controlling for sector, time and state fixed effects will remove variation common to firms across sectors (including sector-specific productivity trajectories shocks), states and period (including aggregate demand shocks). Nevertheless, correlated shocks may still bias my estimates in spite of these adjustments. In the next two sub-sections, I introduce two novel instrumental variable strategies to deal with this concern.

4.3. **Historical SSIV instrument.** I construct a historical SSIV instrument that allows me to estimate the causal impact of spillovers from public R&D on firm productivity from 1950 to 2020. This instrument has the advantage of covering a long time period. However, it cannot be used to estimate the causal impact of private spillovers on firm outcomes, a weakness my second instrument addresses.

The instrument combines agency-specific shocks in federal funding and the shares of exposure to knowledge spillovers s_{iat}. The shocks come from variation in total R&D outlays by 17 government agencies and departments (henceforth, just 'agencies') who have funded some patented innovations, over 13 five-year periods, from 1950 to 2010. Following the notation of equation (3), agencies are indexed by a and periods by t. The identification thus relies on cross-sectional *and* time-variation in agencies' budgets. They consist of the following departments and agencies, in decreasing order of patenting activity in 2010: the Department of Defense (including DARPA), the Department of Health and Human Services (including the National Institutes of Health), the Department of Energy (including ARPA-E), the National Science Foundation, NASA, the Department of Agriculture, the Department of Commerce, the Small Business Administration (including its SBIR seed fund for innovative startups), the Department of Veterans Affairs, the Department of Education, the Environmental Protection Agency, the Department of Transportation, the Department of Homeland Security, the Department of Interior, the Atomic Energy Commission and the Department of State.[28]

To better understand where the variation used in my identification come from, panels A.7, A.8 and A.9 in the Appendix show time series of the budgets of selected agencies. The figures suggest that there is a large degree of heterogeneity and stochasticity in budget changes across agencies and over time. Moreover, a lot of the variation is driven by political decisions or geopolitical events that are plausibly uncorrelated with firm performance and innovation five to ten years later, unless perhaps through spillovers. For instance, changes in spending patterns by the Department of

[28]Some agencies do not exist over the whole 1950-2010 period (*e.g.* NASA, NSF). In periods when an agency does not exist, the shares s_{at} are equal to 0 and the sum of shares for other agencies are equal to 1.

Defense, NASA, the Department of Energy and the Department of Homeland Security are clearly the result of wars, foreign threats, space races, terrorist attacks, the oil shock and other geopolitical events. These are some of the most active agencies when it comes to filing patents and firms are therefore largely exposed to these agencies' innovations. Even agencies without a clear strategic or political mission are subject to variations in funding driven by political events. The National Science Foundation for example, experiences a sluggish budget growth during the Korea war as resources are directed toward the war effort. Conversely, its large budget increase that started in the late 1950s is the result of specific laws triggered by the successful launch of Sputnik in 1957. Similarly, the 1983 increase is due to a sudden decision by the Reagan administration to increase funding for science and engineering.[29] To summarize, changes in federal agencies' budget offer pausibly random variation that is uncorrelated with firm outcomes. In robustness checks, I also use only a subset of funding shocks that are most evidently random based on my read of the agencies histories and the classification of narrative shocks by Fieldhouse and Mertens (2023). This approach can be seen as a 'narrative-SSIV' (more details are provided in 5.1).

The funding shocks are calculated as the log yearly R&D budgets of agencies, deflated using the Bureau of Labor Statistics CPI,[30] and measured at $t-5$, five years before the agencies' patents. The funding shocks are denoted by g_{at}.

$$g_{at} := \ln(\text{R\&D budget}_{at-5}) \tag{6}$$

These shocks are used to construct the firm-specific instrument, $\sum_a s_{iat} g_{at}$, for the endogenous exposure to public R&D spillovers, $\sum_a s_{iat} p_{at}$. Equation (5) is then estimated by Two-Stage Least Squares (2SLS). The endogenous exposure to private R&D spillovers is not instrumented in the SSIV setting.

Out of a theoretical maximum of 208 shocks ($|A| \times |T| = 13 \times 17 = 221$), 172 are used in my empirical exercise because some agencies did not exist for the full period over which I observe firm outcomes and, in some rare occasions, the absence of technological overlap between firms and some agencies in some periods. The quasi-experimental SSIV design relies on numerous, uncorrelated and as-good-as-random shocks. To check if shocks are numerous enough and not dominated by one agency $\times$ period, I compute the inverse of the Herfindahl index of average exposure shares at the level of the identifying variation. A high value of the HHI indicates a dispersed source of variation across agencies and periods and is a necessary condition for the

[29]For a detailed history of the NSF, see 'The National Science Foundation: A Brief History' (1994), by George T. Mazuzan https://www.nsf.gov/about/history/nsf50/nsf8816.jsp. Retrieved January 2023.

[30]Amounts are expressed in 2020 dollars, using the BLS CPI series CUUR0000SA0: data.bls.gov/timeseries/CUUR0000SA0).

consistency of the SSIV estimator and the asymptotic validity of the exposure-robust confidence intervals (Borusyak *et al.*, 2022). Formally, I calculate:

$$\text{inverse HHI} := \frac{1}{\sum_{a,t}^{A,T} s_{at}^2} \qquad \text{where} \qquad s_{at} := \frac{1}{N_{at}} \sum_i s_{ait} \tag{7}$$

that is, I compute the inverse HHI of average shares of exposures of firms, indexed by i, exposed to a in t.[31] Average shares of exposure s_{at} are calculated over all N_{at} firms exposed to agency a at t. The inverse HHI in my sample is 138, suggesting a reasonably dispersed set of shocks.[32] For inference, this value is well above threshold of 20 at which exposure-robust standard errors are close to their asymptotic counterparts (Borusyak *et al.* 2022, p. 199).

The highest such shares of exposure are informative about the variation I am using; they show to which agencies, in which periods, firms in my sample are most exposed. The highest 6 shares are all associated with NASA or the Department of Defense in the late 1950s to early 1970s, consistent with the importance of these two agencies in federal R&D funding in this period. The department of Health and Human Services, the department of Energy and the department of Agriculture in the 1960s and 1970s are completing the top 10.[33] Along with a strong, exposure-robust, first stage F-stat and an absence of pre-trends (both discussed in section 5), the high inverse HHI is indicative of the appropriateness of the SSIV design.

4.4. **Patent examiner leniency instrument.** While the historical SSIV setting enables me to estimate γ–the impact of public R&D spillovers on firm productivity growth–exogenous shocks in agencies' budgets cannot be used to estimate ε, the impact of private R&D. In this section, I present another quasi-experimental identification strategy that addresses this limitation. It relies on patent examiners' leniency, defined as their rate of conversion of patent applications into patent grants, and it enables me to compare the magnitude of spillovers from public agencies to that of spillovers from private firms. The drawback of this approach is to not be applicable to the whole period over which I observe firm outcomes. The patent application data which is used to calculate examiners' leniency is indeed only available from 2001 onward. The results of this approach are therefore complements and not substitutes to the historical SSIV results. I describe this identification strategy in more details in this sub-section.

[31] I use Borusyak *et al.* (2022)'s command to transform my dataset at the firm × period level into a dataset at the level of the identifying variation (agency × period), with corresponding exposure weights.

[32] If one were to run the SSIV specification at the level of agencies × period, like in the Borusyak *et al.* (2022) setting, this would means that the effective sample size used is 103.

[33] The order is as follows: NASA-1970 (2.8%), Defense-1970 (2.6%), Defense-1965 (2.3%), NASA-1965 (2.0%), Defense-1960 (1.9%), Defense-1955 (1.6%), HHS-1970 (1.6%), Energy-1970 (1.5%), HHS-1965 (1.4%) and Agriculture-1965 (1.3%).

Examiners all have the same mandate: grant patents to inventions that are non-obvious, novel and useful. In practice however, they have some discretion when deciding to grant a patent. Examiners vary greatly in their average grant rate, even within years and within the narrow technological categories within which they officiate ('art units', which are different from patent classes). The leniency of an examiner, in turn, has a strong positive association with the probability a patent application is converted to a patent grant.

Previous work has showed that assignment of applications to examiners can be treated as random, conditional on years $\times$ art unit fixed effects (Sampat and Williams, 2019; Farre-Mensa *et al.*, 2020). The random allocation of applications to examiners of varying leniency therefore provides interesting quasi-experimental variation in patent grants, which can be used to study the impact of being awarded a patent on firm outcomes. The innovation literature has made extensive use of this 'patent lottery' (Farre-Mensa *et al.*, 2020) to study, among other, patent litigation (Feng and Jaravel, 2020), startup growth (Farre-Mensa *et al.*, 2020) and, like in the present context, spillovers (Sampat and Williams, 2019). In my setting, I am using examiners' leniency in a novel way: not at the level of the focal firm whose outcomes I am interested in, but at the level of the agencies a focal firm is drawing inspiration from.

The patent lottery is used here to affect spillovers. Some firms happen to be exposed to spillovers by entities who were fortunate to face more lenient examiners. Other firms are receiving fewer spillovers because upstream patent examiners were more conservative. The patent examiner instrument acts as a randomizing device for upstream patent generation, conditional on a suitable set of covariates. It therefore approximates the ideal experiment of randomizing knowledge production by agencies and firms.

The identification relies on the creation of an instrument for $\sum_a s_{iat} p_{at}$ and $\sum_f s_{ift} p_{ft}$, the exposures to patent production by agencies and firms. The instruments are weighted average leniencies faced by upstream agencies $\sum_a s_{ia} \bar{l}_a$, and by upstream firms $\sum_f s_{if} \bar{l}_f$. In both instruments, the shares are calculated like in the historical shift-share instruments using (4). Average leniencies are calculated as $\overline{l_{a,t}} = \sum_{j \in J_{at}} \frac{l_{e(j),t}}{|J_{at}|}$: the average of examiners' leniencies $l_{e(j)t}$ across the set of all the applications that agency a submits in year t. This set is denoted J_{at}. Applications are indexed by j and examiners by e. Examiner leniencies for an agency are calculated using all applications submitted to an examiner, excluding those submitted by the agency in question. This creates leave-one-out leniency indices that are agency-specific. They are further residualized on art units and years. The exposure to average leniency of upstream agencies $\sum_a s_{ia} \bar{l}_a$ can then be used as an instrument for the change in exposure to spillovers by these same upstream agencies $\sum_a s_{ia} p_a$. The next section shows that this instrument is strong for both private and public R&D. As for the exclusion restriction , it is likely to be satisfied due to the quasi-experimental nature of the allocation of applications to examiners.

Discussion. What are the mechanisms through which the instrument work? There are two potential mechanisms. The first is the validation of the quality of an innovation. An innovation protected by a granted patent is more likely to be of higher quality than a non-granted innovation because it satisfies the criteria of usefulness, non-obviousness and novelty used by patent examiners to grant patents. This makes the granted patent a more powerful vehicle for spillover because of this 'seal of approval' from the USPTO. The second mechanism is the revelation of the innovation to the wider world. Patent applications are confidential for 18 months from the date of filing. This so-called 'pendency' lag covers almost entirely the average lag between patent applications and grant that USPTO patent applicants have historically experienced (around 20 months). Patents that are granted before the 18 months of secrecy therefore provide a visibility boost to their innovation, in addition to the signal of quality. Moreover, patent applicants can decide to opt out of the automatic disclosure of application. Over the period covered by my instrument (2000-2010), around 10% of applicants opt out when applying.

One concern about the validity of this IV approach is that aggregating leniency scores of examiners across all the applications of an agency will lead to a lack of usable variation in the instrument. Agencies indeed draw successive, plausibly independent and random examiner leniencies when they submit several patent applications. The average examiner leniency they are exposed to will therefore converge in probability to 0–the population mean of leniency scores residualized on art units $\times$ year–as their number of applications grow, by the Law of Large Number. The larger the volume of application an agency files, the smaller the variation in average leniency scores. This may then lead to a weak first stage and and invalidate this IV design. The problem may be more severe for the public R&D instrument because public agencies have typically higher volumes of applications than firms.

To mitigate this concern, I define public agencies as the actual assignees and/or funding agencies of patents as reported in the USPTO data, rather than aggregating public agencies at the coarse level for which I have data on R&D budgets like in the SSIV design. Patent applications are therefore linked to entities such as the Lawrence Livermore National Laboratory or the Advanced Research Projects Agency–Energy (ARPA-E) rather than the wider Department of Energy to which they belong. There are 200 such fine agencies compared to the 17 used in the historical SSIV. This step reduces the average volume of agencies' applications and thus mitigates the risk of the variation in average leniencies to collapse to 0. Figure 18 in the Appendix shows that this step leaves a lot of useful variation in the average leniencies faced by agencies and firms, if they file fewer than 20 patent applications. In my data, 90% of firms and 60% of fine agencies file fewer than 20 patents a year. Shares of exposure to spillovers are appropriately calculated over these 200 fine agencies and thousands of private patent assignees.

5. RESULTS

I now turn to the regression results from the two instrumental variable strategies, starting with the historical SSIV.

5.1. Historical SSIV. My main sample consists of 6,499 firm-by-period observations for which outcome variables, pre-trend outcomes and controls are not missing. Firms in 'Finance, Insurance and Real Estate' are excluded. Observations are further selected on non-missing exposures to public or private spillovers. Table E.16 in the Appendix provides summary statistics about the sample. Firms are rather large, with a median employment count of 5,000 workers, median yearly sales of 1.2 billion 2020 USD and 4 million in yearly median R&D expenses. Filing patents in any given year is relatively rare; the median firm files three. The most represented sectors are in electronic components, lab apparatus & instruments, and surgical, medical, & dental instruments and supplies.

For all SSIV results, standard errors are robust to arbitrary correlation across firms that are exposed to a similar distribution of agencies, using the method developed by Adão *et al.* (2019). Adão *et al.* (2019) show that clustered or heteroskedasticity-robust standard errors may substantially underestimate the variability of IV estimators when the instrument takes a shift-share form. The reason is that the regression residual v_{it} in (5) will include shift-share-like terms with shares correlated with the shift-share instrument. This leads firms with similar exposure shares to have similar exposures to the shocks and then similar residuals. This correlation structure is likely to exist in my setting: firms more exposed to innovation by NASA, for instance, may have correlated productivity growths that standard errors clustered at the sector or state level may fail to account for.

First stage. The validity of the SSIV identification relies on a strong first stage *i.e.* a strong relationship between funding shocks at $t-5$ and patent production funded by these agencies at t. Figures 4a and 4b provide evidence that such a relationship exists. Figure 4a shows a scatterplot of the public R&D spillovers variable, $\sum_a s_{iat} p_{at}$, residualized on sector, period and state fixed effects as well as lagged firm controls (R&D, employment, capital and patent count) on the average of R&D

funding shocks, $\sum_a s_{iat} g_{at}$, also residualized. The relationship is positive and significant, with an exposure-robust F-stat of 98, suggesting that the instrument is strong.[34]

To gauge the appropriateness of the timing, and in particular the five-year lag separating funding shocks to patent production by agencies, Figure 4b provides a visual assessment of the dynamic relationship between the two by reporting the impulse response of patents to R&D funding at various time horizons. It reports point estimates and confidence intervals of local projections of yearly patent production by federal agency (in log patents) on R&D funding levels (in log 2020 dollars), where patent production is observed at different years relative to the funding. The specification controls for year and agency fixed effects, and for five lags of funding.[35] The regressions are weighted by patent counts at time $t = 0$ to account for the greater importance of large agencies in the composition of federal R&D, and thus in the shares of exposures of firms to federal innovation. Newey-West standard errors (Heteroskedasticity and Autocorrelation Consistent) with one lag are reported (95% and 90% levels). The figure shows that an agency's patents production after a funding shock is positively associated with the (log) amount of funding at t. The elasticity progressively increases after the funding shock, until it reaches a maximum of 0.45 at $t + 9$ before slowly coming back down to its baseline level. While the impulse response is imprecisely estimated, patent production clearly shows an upward trend after the shock. Interestingly, patent production before the funding shock does not appear to be correlated with the shock. This provides some evidence that the R&D funding variation that I exploit is not a consequence of underlying productivity or innovation trends (captured by agencies' patent productions).

The delay between public funding of R&D and patent production is in line with the evidence reported in previous research. De Rassenfosse *et al.* (2019) find that the average gestation lag between a US government procurement contract being awarded to a firm and the filing of a patent by this firm is 33 months (2.75 years), with 90% of all patents linked to contracts being filed between 1 and 7.5 years.[36] Azoulay *et al.* (2019) study grants from the NIH to pharmaceutical firms and find longer delays: two thirds of grantees who eventually file a patent, file it within 10 years

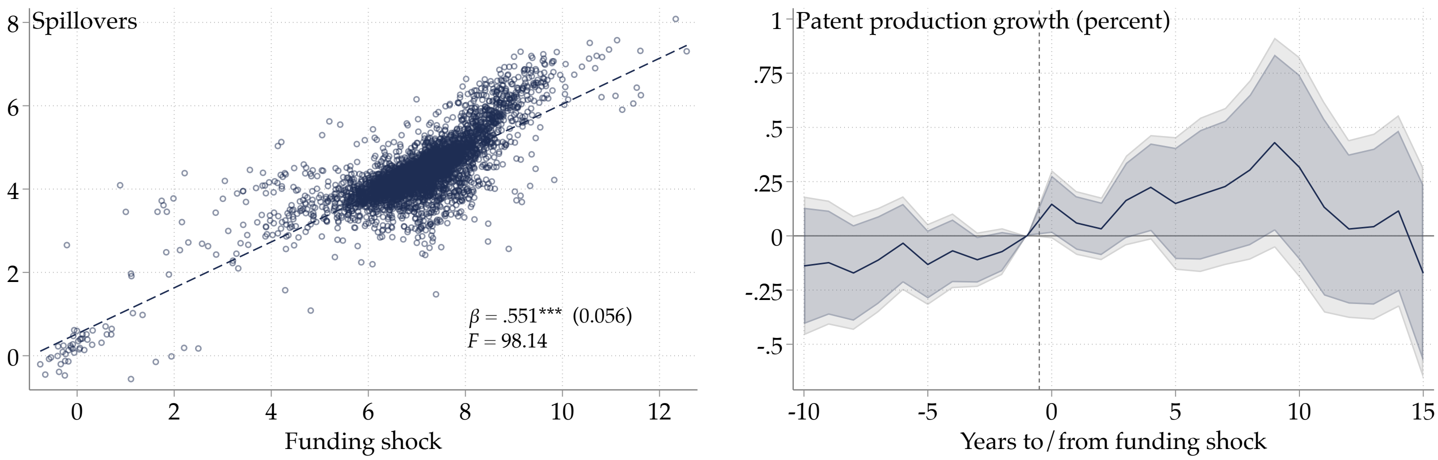
[34] The corresponding sector-clustered Cragg-Donald F-stat is 89.4. The lower value of the exposure-robust F-stat highlights the relevance of exposure robust inference in my setting.

[35] For a given lag τ, the estimating equation is:

$$p_{a,t+\tau} = \beta x_{at} + \gamma' \boldsymbol{X}_{at} + \delta_i + \tau_t + \epsilon_{it}$$

$p_{a,t+\tau}$ is the log count of patents by agency a in year $t + \tau$, x_{at} is the log R&D budget of agency a in focal year t and the vector $\boldsymbol{X}_{at}$ contains lags of R&D budgets. The coefficient of interest is β. τ varies from -10 to +15.

[36] Own calculations based on figure 2A of De Rassenfosse *et al.* (2019).



(A) First-stage graph: SSIV

(B) Local projection of patent production

FIGURE 4. SSIV first stage

Notes: The left-hand side scatterplot shows the correlation between public R&D spillovers and exposure to funding shocks for all 6,499 firms × period in my historical SSIV sample. Standard errors and first-stage F-stats are exposure-robust (Adão *et al.*, 2019). Spillovers and exposure to funding are partialled out on the full set of controls used in (5).
The right-hand side graph is a local projection of (log) patents by federal agencies on their (log) R&D funding, at different time horizons. The unit of analysis is a federal agency ($N = 17$). Standard errors are Heteroskedasticity and Autocorrelation Consistent (Newey-West with one lag).

of the award data and nearly all firms who do file a patent do it within 15 years.[37] Overall, the empirical exercise of Figure 4b and previous research provide supporting evidence for the timeline described in Figure 3.

Main impacts. I report here on three sets of 2SLS regression results, all using a stacked difference specification that divides the 1950-2020 panel into equally sized 5-year intervals to estimate equation (5). The first set of results, shown in Table 1, reports changes in productivity, sales and employment from t to $t+5$, and flow variables such as patent production and R&D investments at $t+5$. I further report the probability of filing a patent at $t+5$ to evaluate the extensive margin impact of public R&D spillovers on innovation. In all specifications, standard errors are exposure-robust (Adão *et al.*, 2019). To investigate the sensitivity of my 2SLS estimates, I report coefficients $\widehat{\gamma}$ across specifications with increasingly comprehensive controls. All specifications include sector, period and state fixed effects. To investigate the importance of the coarseness of sector fixed effects specifically, I present in the last columns coefficients obtained when controlling for 238 fine 3-digit sectors (like '382 – Measuring and Controlling Devices') instead of the 65 coarser 2-digit sector fixed effects (like '38 – Instruments and related products'). Starting from the simplest specification, including only own R&D effort, patents, and period, state and sector fixed effects, in column

[37] Figure 5, p. 135.

(1), I progressively add the endogenous private R&D spillovers in (2) and lagged firms' capital, employment and sales in (3). Importantly, all first stage F-stats are high; they hover at around 100.

Overall, the results shown in Table 1 suggest that an increase in exposure to public R&D spillovers has a positive impact on a broad range of firm-level productivity indicators and own R&D expenditures. It is however notable that firms do not appear to grow more in terms of sales or employment. Coefficients are stable across specifications, even when switching from coarse to fine sector fixed effects.

Productivity at the firm level is estimated using the Olley and Pakes (1996) method with the correction suggested by Ackerberg *et al.* (2015).[38] Using this measure of productivity as my main outcome of interest, I find that a 1% increase in public spillovers causes a .023% to 0.025% increase in productivity across specifications (first rowm of Table 1). Estimated measures of productivity are also positively impacted: Cobb-Douglas and translog productivities are all positively impacted with elasticities between .03 and .04 (significant at the 1% level).

Turning to innovation outcomes, I find that public R&D spillovers positively impact a firm's investment in R&D. Each 1% increase in spillovers cause a .023 to .026% increase in own R&D spending, five years after the shock (penultimate row of Table 1). This result echoes the finding of Moretti *et al.* (2019) who show that public and private R&D are complements: an increase in public R&D tends to *crowd in* private investment in R&D. It also complements the findings of Fieldhouse and Mertens (2023) that R&D appropriation for both defense and non-defense shocks cause private R&D investments to increase.

The impact of public R&D spillovers on innovation by the focal firm is also notable. To deal with the large number of zeroes in the patent count field, I use the Inverse Hyperbolic Sine of patent counts at time $t+5$ rather than the log of patents.[39]. It appears that firms increase their own patent production following a positive spillover shock: each 1% increase in spillovers generates a more than 0.02% increase in own patent production. Finally, the last column of Table **??** shows that public R&D spillovers also impact a firm's propensity to file patents five years down the road.

Pre-trends and falsifications. To evaluate the validity of the historical SSIV setting, I conduct falsification tests where I investigate if firms who are more intensively treated were on different growth trajectories before time t. To do so, I regress lagged outcomes (measured from $t-5$ to t, or at t for flow variables) on the instrumented exposure to spillovers and the suite of controls of specification (5). Results are reported in Table 2. I find that firms more exposed to spillovers do not appear to be on a significantly different trajectory than firms less intensively treated. In the fullest

[38]COGS are used as variable inputs, the state variable is the capital stock (PPEGT) and the instrument is investment (CAPX). Estimation is performed with Stata's `prodest` package.

[39]$\text{IHS}(x) = \ln\left(\frac{x}{2} + \frac{1}{2}\sqrt{x^2+1}\right)$. The Inverse Hyperbolic Sine behaves like the natural logarithm for large values of x, but it is defined at $x = 0$.

	(1)	(2)	(3)	(4)
Productivity				
$\Delta_5 \ln(\text{TFP})_t$	.024**	.025***	.025***	.023**
	(.009)	(.009)	(.009)	(.011)
Firm sales and employment				
$\Delta_5 \ln(\text{Sales})_t$	.009	.009	.008	.010
	(.008)	(.008)	(.008)	(.007)
$\Delta_5 \ln(\text{Employment})_t$	.007	.008	.008	.009
	(.009)	(.009)	(.009)	(.008)
Innovation				
IHS Patent count$_{t+5}$	.021***	.023***	.024***	.026***
	(.007)	(.008)	(.007)	(.009)
$\ln(\text{R\&D})_{t+5}$	.040***	.029**	.031**	.035***
	(.015)	(.013)	(.013)	(.009)
$\Pr(\text{Files patents})_{t+5}$	.016*	.018*	.019**	.017**
	(.009)	(.009)	(.009)	(.008)
First-stage F-stat (exp. robust)	97.34	97.40	98.14	108.14
Period FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
Sectors FE (2-digit)	✓	✓	✓	
Sectors FE (3-digit)				✓
Own R&D and patents	✓	✓	✓	✓
Private R&D spillovers		✓	✓	✓
Lagged firm controls			✓	✓
N	6,499	6,499	6,499	6,499

TABLE 1. Historical SSIV regression results – 5 years

Notes: The unit of observation is a firm × period. This table shows the estimates for ϵ, the impact of a 1% increase in spillovers from public R&D on various firm outcomes (listed in the leftmost column). Standard errors and F-stats are exposure-robust (Adão *et al.*, 2019): they are computed using the authors' `reg_ss` and `ivreg_ss` commands. Lagged firm controls include capital, employment and patent counts (all in logs).
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.

specification (column 3), the coefficient on public R&D spillovers is never significantly different from 0. Most importantly, firm TFP does not exhibit any pre-trend. Some pre-trends appear when I control for fine sectors (column 4); firms experiencing larger increases in public R&D spillovers tend to invest more in R&D already at time t, but they file fewer patents. One explanation for the positive response of time-t R&D can be that private R&D responds more to public R&D funding (at time $t-5$) than to public R&D patent production (measured at time t).

Overall, the absence of pre-trends in my main specification (column 3) provides some credibility to the SSIV setting by ensuring that the positive productivity impacts documented in Table 1 are not a reflection of an already existing positive increase in productivity and innovativeness that would have happened irrespective of the treatment.

10-year outcomes. In Appendix E.2, I use the same sample of firms to test if the productivity increase that happens after 5 years persists over longer horizons. The increase in firm TFP is indeed persistent after 10 years (+.027%, significant at the 5% level), suggesting that firms experience a durable rise in productivity following a one-time spillover shock. Interestingly, a greater exposure to public R&D spillovers cause a slight but detectable *reduction* in employment after 10 years. In other words, firms benefit from technology spillovers by becoming more productive and by economizing on labour, over long-enough durations.

Narrative approach. If R&D expenditures by federal agencies are *reacting* to factors affecting productivity trends, the quasi-experimental SSIV approach I am using may not be appropriate. My estimates would then capture a (plausibly positive) correlation between investments by federal agencies in certain technologies and the upward productivity growth of firms who are active in the use or development of these technologies. The absence of pre-trends documented in Table 2 provides some evidence that this issue is unlikely to be present in my setting. Nevertheless, I provide further validation for my quasi-experimental approach by selecting agency funding shocks that are likely to be uncorrelated with other factors affecting productivity trends. This narrative approach is similar to that of Fieldhouse and Mertens (2023) and I partly rely on their selection of historical funding shocks to select mine. I further add shocks experienced by the National Science Foundation and the department of Homeland Security to my list of narrative shocks. The shocks I keep in my narrative-SSIV are listed in Tables E.18 and E.19 in Appendix E.3, along with a justification for their inclusion. This procedure gives me a list of 47 shocks. The Department of Defense is the most represented agency among these shocks (15 shocks in total). It is followed by the National Science Foundation (9) whose funding is eminently political. For instance, its research priorities in the 1950s were set by the urge to keep a technological lead over the USSR, and the NSF is usually one of the first agencies to get its funding reduced in times of tight budget controls, like after the Budget Control Act of 2011.

Figures 5a and 5b show how the estimates from the narrative-SSIV approach compare to those of the standard SSIV for the main productivity outcomes I am interested in. First, it is notable that the exposure-robust F-stat is slightly lower when using the narrative-SSIV; its value is 48.25 compared to 98.14 (column 3 of Table 1): the narrative-SSIV instrument uses less variation than what is

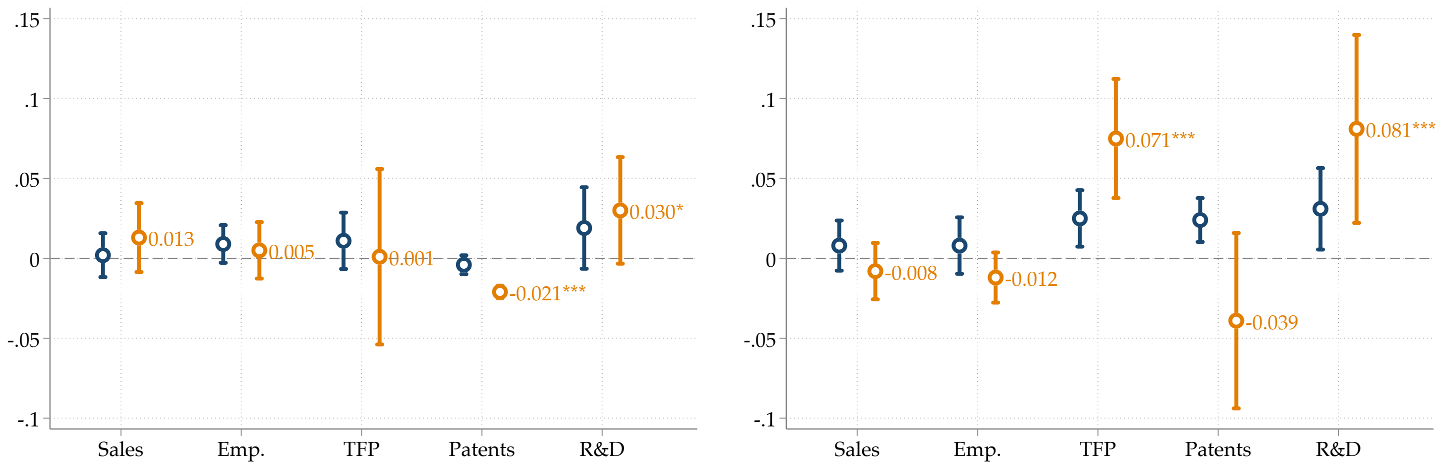
	(1)	(2)	(3)	(4)
Productivity				
$\Delta_5 \ln(\text{TFP})_{t-5}$	.014	.011	.011	.014
	(.009)	(.009)	(.009)	(.011)
Firm sales and employment				
$\Delta_5 \ln(\text{Sales})_{t-5}$	.003	.003	.002	.004
	(.008)	(.008)	(.007)	(.007)
$\Delta_5 \ln(\text{Employment})_{t-5}$	.009	.01	.009	.011*
	(.006)	(.006)	(.006)	(.006)
Innovation				
IHS Patent count$_t$	-.005*	-.004	-.004	-.005**
	(.003)	(.003)	(.003)	(.003)
$\ln(\text{R\&D})_t$	.026*	.018	.019	.021**
	(.014)	(.013)	(.013)	(.009)
$\Pr(\text{Files patents})_t$	-.003	.000	.000	-.003
	(.009)	(.010)	(.010)	(.009)
First-stage F-stat (exp. robust)	97.34	97.40	98.14	108.14
Period FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
Sectors FE (2-digit)	✓	✓	✓	
Sectors FE (3-digit)				✓
Own R&D and patents	✓	✓	✓	✓
Private R&D spillovers		✓	✓	✓
Lagged firm controls			✓	✓
N	6,499	6,499	6,499	6,499

TABLE 2. Historical SSIV regression results – Pre-trend tests

Notes: The unit of observation is a firm × period. Standard errors and F-stats are exposure-robust (Adão *et al.*, 2019): they are computed using the authors' `reg_ss` and `ivreg_ss` commands.
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.

available across the intersection of agencies and time periods and this results in a slightly weaker first stage. The second stage results are however broadly similar across the two specifications. The narrative-SSIV coefficients indicate no pre-trend across most outcomes. However, patent production is significantly negative the pre-period when using the narrative SSIV. Turning to 5-year firm outcomes, nearly all narrative-SSIV coefficient are very close to the SSIV ones with the exception of the coefficient on spillovers when patent production is the dependent variable; the coefficient is indistinguishable from 0 in this specification. Overall, the narrative-SSIV approach provides some support for the quasi-experimental SSIV approach. With the exception of the specification when patents are on the right-hand side, restricting shocks to those that are evidently exogenous does not affect the results much.

It is also notable that the elasticity of productivity to public R&D spillovers is higher (+.071%) when using the exogenous shocks than when using all shocks (+.025%). This result is a likely manifestation of heterogeneous impacts across federal agencies.



(A) Pre-trends

(B) 5-year outcomes

FIGURE 5. Comparison of the SSIV (blue) and narrative-SSIV (orange)

Notes: The figures show point estimates and 95% confidence intervals of the coefficients of exposure to spillovers, instrumented by the SSIV (in blue) and narrative-SSIV instruments (in orange). Estimates come from my preferred specification of column (4) in the regression tables. The unit of observation is a firm $\times$ period. Standard errors and *F*-stats are exposure-robust (Adão *et al.*, 2019): they are computed using the authors' `reg_ss` and `ivreg_ss` commands.
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.

Treatment heterogeneity. The discussion so far has postulated a constant causal effect of spillovers on firm growth, across all firms. I present here estimates of treatment effect heterogeneity by firm size.

Several reasons motivate the focus on treatment heterogeneity. Firstly, there has been a secular trend toward more concentration among American businesses, in particular since the 1960s, as documented by Kwon *et al.* (2022). Research into the causes of the rise in concentration is still very active and of prime policy interest. Previous work has emphasized the role of technology (Autor *et al.*, 2020; Hsieh and Rossi-Hansberg, 2023), a lack of competition, perhaps caused by a lack of appropriate regulation (Gutiérrez and Philippon, 2017), increased barriers to entry (Furman, 2015), decreasing spillovers between market leaders and followers (Akcigit and Ates, 2022;

Olmstead-Rumsey, 2022) or globalization (Feenstra and Weinstein, 2017). My empirical exercise suggests another, complementary explanations: smaller firms rely more on spillovers from public R&D than larger firms and the decline in public R&D might therefore put smaller firms at a disadvantage.

Secondly, as fact 3 in section 3 showed, smaller firms are more likely to cite public R&D patents which points to he importance of spillovers for them. Prior work has shown that firms of different sizes use spillovers differently. Acs *et al.* (1994) for instance, were the first to document that smaller US firms make a more extensive use of spillovers than large ones. By contrast, large corporations rely more on their own R&D investments. The theoretical argument is that, with a lesser capacity to mobilise own R&D funds, small firms tend to rely on another complementary input in their knowledge production function: ideas from other sources. Audretsch and Vivarelli (1996) finds similar results among Italian firms.

To test if smaller firms in my data benefit more from public R&D spillovers than larger ones, I modify my estimating equation (5) by adding the interaction of the public spillover variable with the natural log of firm employment in $t-5$, taken here to represent firm size. I demean firm size by average log employment. The coefficient on the interaction term can thus be interpreted as the marginal impact of a 1% increase in spillovers on the productivity of a firm that is one log-point larger than average. At the average firm size of 23,000 employees, this one log-point difference corresponds to a jump to 62,500 employees (an e^1-fold increase). Equivalently, this is comparable to the difference between the median firm (5,000 employees) and a firm at the 70$^{\text{th}}$ percentile (13,5000). The estimating equation for the interaction effect is:

$$\begin{aligned}\Delta z_{it} =& \phi e_{it} + \gamma_1 \sum_a s_{iat} p_{at} + \gamma_2 \sum_a s_{iat} p_{at} \times \ln\left(\widetilde{\text{emp}_{it-5}}\right) \\ &+ \varepsilon \sum_f s_{ift} p_{ft} + \eta_{s(i)} + \tau_t + \lambda_{g(i)} + \mathbf{X}_{it}\boldsymbol{\beta} + v_{it}\end{aligned} \tag{8}$$

where γ_1 is the baseline impact and γ_2 is the interaction effect. $\ln\left(\widetilde{\text{emp}_{it-5}}\right)$ stands for demeaned employment at $t-5$. Public R&D spillovers and their interaction with size are instrumented by funding shocks and funding shocks interacted with size, respectively. Standard errors are exposure-robust. As shown in Table 3, heterogeneity of the impact of spillover matters, and the coefficients on the treatment interacted by firm size have a negative sign for productivity, sales and employment: larger firms are less likely to benefit from spillovers from public R&D along these dimensions. The baseline impact on TFP is positive, suggesting that all firms benefit from spillovers. Baseline elasticities .018% to .023%, in line with the main impacts found in Table 1. This positive effect on productivity is quickly decreasing with firm size though; a firm one log point larger than its peers experiences a .006% lower increase in value added per worker due to public

R&D spillovers, as can be seen from the point estimate of γ_2 in column 3 of Table 3. Taken at face value, and assuming that the log-linear relationship between spillovers and firm size holds further away from the average firm size, this means that a firm 3.6 log-point bigger than the average firm experiences no productivity growth from public R&D spillovers. While firm sales and employment did not appear to be affected by public R&D spillovers in the baseline specification, small firms experience large gains in size according to the coefficients on the interaction term reported in table 3. A firm 1-log point smaller than the average firm grows by .01% (.0035+.0071, column 3) in terms of sales and by .013% (-.0021+.0146, column 3) in terms of employment count.

Interestingly, larger firms are more likely to file patents following an increase in public R&D spillovers. This finding points to the greater reliance of large firms on the patent system to protect their IP (Mezzanotti and Simcoe, 2023). They are also investing in R&D at a higher rate than smaller firms.

Summary and discussion. This section has reported on several empirical exercises using a historical SSIV identification to identify the causal impact of public R&D spillovers on firm productivity. I have documented that a 1% larger public R&D spillover shocks translate into .025% higher productivity (TFP estimated via the Olley and Pakes (1996) methodology) at the firm level. I have also shown that small firms are benefiting much more from these spillovers when it comes to productivity, sales and employment growth. One drawback of the SSIV approach is that I cannot compare the magnitude of impact of public spillovers to that of private spillovers. The next sub-section turns to my second instrument to make progress on this front.

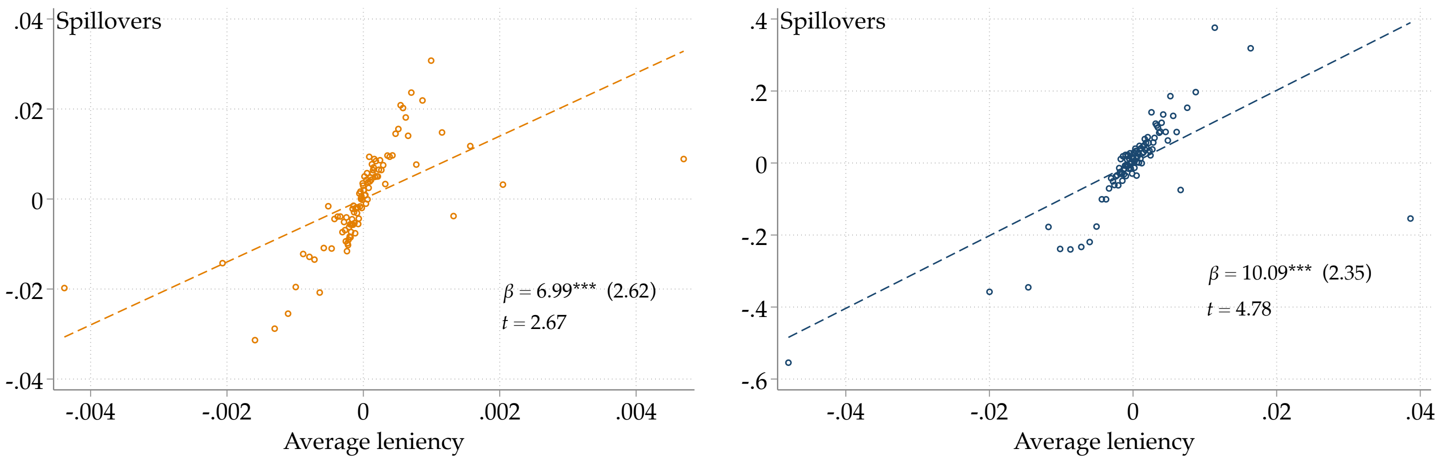
5.2. **Patent examiners regressions.** Patent examiner regressions provide interesting evidence that spillovers from public agencies are between two and three time as impactful as spillovers from the private sector when it comes to increasing private firms' productivities.

Examiner leniency instrument first stage. For both the public and private R&D instrument, the first stage is rather strong, with F-statistics around 18 and 6.4, respectively, as can be seen in figure 6 which plots the endogenous exposure to spillovers as function of the exogenous instrument using examiners' leniency, for the private and public exposures to spillovers. Both quantities are partialled out on the set of controls used in the regression results. The joint F-stat (Cragg-Donald) is 56.3 for my main specification. Because the identifying variation in my patent examiner regressions come from the examiners and not the upstream firms or agencies filing patents, exposure-robust F-stats and standard errors are not indicated. I therefore use clustered standard errors at the period $\times$ sector level.

		(1)	(2)	(3)	(4)
$\Delta_5 \ln(\text{TFP})$	*Baseline*	.0205**	.0225**	.0214**	.0183*
		(.0092)	(.0093)	(.0092)	(.0107)
	Interaction	-.0037***	-.0039***	-.0059***	-.007***
		(.0001)	(.0001)	(.0004)	(.0003)
$\Delta_5 \ln(\text{Sales})$	*Baseline*	-.0009	.001	.0035	.0058
		(.0071)	(.0073)	(.008)	(.0072)
	Interaction	-.0114***	-.0117***	-.0071***	-.0059***
		(.0002)	(.0002)	(.0004)	(.0003)
$\Delta_5 \ln(\text{Emp.})$	*Baseline*	-.0052	-.0023	-.0021	-.002
		(.0079)	(.008)	(.0082)	(.0074)
	Interaction	-.0148***	-.0152***	-.0146***	-.016***
		(.0002)	(.0002)	(.0004)	(.0004)
IHS Patent count$_{t+5}$	*Baseline*	.0368***	.0359***	.0344***	.0354***
		(.0081)	(.0082)	(.0078)	(.0093)
	Interaction	.0188***	.0189***	.016***	.0146***
		(.0003)	(.0002)	(.0005)	(.0005)
$\ln(\text{R\&D})_{t+5}$	*Baseline*	.0613***	.0481***	.0445***	.0524***
		(.0155)	(.0127)	(.0131)	(.0085)
	Interaction	.0252***	.0268***	.02***	.0262***
		(.0006)	(.0003)	(.0007)	(.0005)
$\Pr(\text{Patents})_{t+5}$	*Baseline*	.0176**	.0199**	.0202**	.0189**
		(.0088)	(.0095)	(.0095)	(.0084)
	Interaction	.0022***	.002***	.0027***	.0032***
		(.0002)	(.0002)	(.0004)	(.0004)
First-stage F-stats	*Baseline*	97	98	98	108
(exposure robust)	*Interaction*	>1,000	>1,000	>1,000	>1,000
	Joint[40]	863	905	902	898
Period FE		✓	✓	✓	✓
State FE		✓	✓	✓	✓
Sectors FE (2-digit)		✓	✓	✓	
Sectors FE (3-digit)					✓
Own R&D and patents		✓	✓	✓	✓
Private R&D spillovers			✓	✓	✓
Lagged firm controls				✓	✓
N		6,499	6,499	6,499	6,499

TABLE 3. Historical SSIV regression results – Heterogeneity of impact by firm size

Notes: Standard errors and individual F-stats are exposure-robust (Adão *et al.*, 2019): they are computed using the authors' `reg_ss` and `ivreg_ss` commands.
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.



(A) Private spillovers

(B) Public spillovers

FIGURE 6. First stages

Notes: The graphs show the correlations between the endogenous treatment, $\sum_a \ln(\text{patents}_a)$ (the average exposure to spillovers from agencies or firms indexed by a), and the instrument, $\sum_a \overline{\text{leniency}_a}$ (the average leniency faced by agencies or firms indexed by a). Both the endogenous treatment and the instruments are residualized on periods, states and 3-digit sectors fixed effects, as well as lagged R&D capital, employment and patent count. This corresponds to specification (3) in Table 4.
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively. Standard errors are clustered at the period and 2-digit sector level.

Patent examiner IV results. In Table 4, I report the results of estimating equation (5) by 2SLS when exposure to public and private spillovers are instrumented by $\sum_a s_{iat}\bar{l}_{at}$ and $\sum_f s_{ift}\bar{l}_{ft}$, respectively, the average leniencies to which upstream patent assignees are exposed to. The sample consists of 5,846 firm×period observations. In line with equation (5), I control for the lagged R&D expenditure of firms to capture increases in own productivity not directly attributable to spillovers, as well as the progressively more exhaustive suite of controls used in the historical SSIV regressions. I present results for my main measure of productivity, as well as a test for pre-trends for this outcome.

The results in Table 4 suggest that firm level productivity increases by more following a shock to public spillovers than after a shock to private spillovers. In my preferred specification with all controls and SIC2 industry fixed effects (column 4), a 1% increase in public spillovers causes a 0.08% increase in productivity (significant at the 1% level). This estimate is not too far from the .07 elasticity that I obtained with the narrative-SSIV specification, but it is higher than the .025 elasticity from the baseline estimate. One tentative explanation for the discrepancy is that the period over which the patent examiner instrument is used (2000-2010) is one of sustained productivity increase in the American economy. The higher impact of public R&D spillovers here might capture some of this effect.

In contrast, a 1% increase in private spillovers causes an increase in productivity of only a third to a half of that amount. But the estimate of the private spillovers coefficient is not statistically different from zero and it is imprecisely estimated. The evidence about the different impacts of public and private R&D spillovers is mixed.

Table 4 also reports pre-trend tests on firm productivity, in the spirit of those reported for the historical SSIV instrument. Across specifications, there is no pre-trend in productivity.

	(1)	(2)	(3)	(4)	(5)
Main outcomes – Dependent variable: $\Delta_5 \ln(\text{TFP})_t$					
Public spillovers	.085**		.082***	.084***	.086***
	(.037)		(.024)	(.023)	(.022)
Private spillovers		-.361	.0303	.0423	.0134
		(.576)	(.226)	(.230)	(.447)
Pre-trends – Dependent variable: $\Delta_5 \ln(\text{TFP})_{t-5}$					
Public spillovers	-.0197		-.0101	-.0101	-.0142
	(.0401)		(.0343)	(.0376)	(.0452)
Private spillovers		.185	.204	.187	.122
		(.653)	(.514)	(.555)	(.727)
First-stage F-stats					
Public spillovers	408		19.1	18.4	16.2
Private spillovers		246	6.4	6.4	6.3
Joint			57.7	56.3	45.9
Period FE	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓
Sectors FE (2-digit)	✓	✓	✓	✓	
Sectors FE (3-digit)					✓
Own R&D and patents	✓	✓	✓	✓	✓
Lagged firm controls				✓	✓
N	5,846	5,846	5,846	5,846	5,846

TABLE 4. Patent examiner regression results

Notes: The unit of analysis is a firm × period. Coefficients and 95% intervals show the results of a 2SLS estimation of (3), where private and public R&D spillovers are instrumented by exposures to changes in average leniencies faced by upstream firms. Lagged firm controls include sales, employment, capital and patent count. ***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively. Standard errors are clustered at the period and 2-digit sector level.

To evaluate if the micro empirical estimates from the historical SSIV and the patent examiner instrument matter for aggregate growth and inequality, I now turn to a general equilibrium model of growth that uses these micro estimates as calibrated parameters.

6. MODEL AND CALIBRATION

Overview of the model. To evaluate the aggregate consequences of the fall in public R&D, I present here a tractable general equilibrium model of growth with heterogeneous firms and spillovers, where public and private R&D are distinct. The theory is inspired by heterogeneous agent models of long-term growth (Luttmer, 2007; Jones and Kim, 2018) and the main theoretical contributions of this paper is to formalize the difference between private and public R&D. This allows me to show how the balance between public and private R&D determines growth and inequality. My model delivers simple, closed-form relationships between the share of researchers funded by the government, aggregate productivity growth and firm inequality.

Unlike in standard endogenous growth models, the central allocative decision does not oppose production to research. Instead, the allocation of funds to basic or applied R&D *within* research determines long-term growth. The strong complementarity between basic (funded by the government) and applied R&D (funded by the private sector) in the generation of spillovers results in a spillover-maximizing split that is interior. Higher spillovers then lead to (i) higher growth through an aggregate boost to all firms and to (ii) lower inequality through easier replacement of incumbents. The main result of the theory (proposition 2) shows that the growth rate follows an inverted-U relationship in the share of basic researchers and so does equality between firms. Consequently, there exists a unique intermediate share of basic researchers that both maximizes BGP growth and minimizes BGP inequality. Current low levels of productivity growth may be due to a share of public R&D that is too low (to the left of the peak of the inverted U).

I calibrate the model from the 1950s onward using the values of elasticities of productivity with respect to public and private R&D estimated in the previous empirical part. The tight link between the model and the estimating equation of section 4 offers a direct mapping from the γ and ε parameters to their quasi-experimentally-estimated counterparts. The calibration exercise suggests that the decline in public R&D matters for aggregate growth and inequality: it explains around a third of the decline in TFP from 1950 to 2017 and a third of the rise in inequality of profits between firms. To save space, proofs and derivations are relegated to Appendix G. Table G.22 summarizes the notation used.

6.1. Firms.

Time is continuous and there are three agents in the economy; researchers (R), workers (L) and firm owners indexed by i, of which there is a unit mass at all times. Total population is fixed and equal to $N = R + L + 1$. Firms' productivity growth is determined by three forces: their

R&D effort, idiosyncratic deviations ('luck'), and an aggregate component capturing the contribution of spillovers to growth. I first present firms' static problem before turning to their dynamic one.

Static firm problem. Each firm produces one variety in a monopolistically competitive environment. Firms' output, denoted y_i, is then aggregated into a final output good via a CES production function. This final output good is the *numéraire* and is equal to GDP (time subscript omitted).

$$Y := \left(\int_0^1 y_i^\theta \mathrm{d}i \right)^{\frac{1}{\theta}} \quad 0 < \theta < 1 \tag{9}$$

where θ is the substitution parameter: a higher value of θ implies an easier substitutability between inputs.[41] A monopolist's production technology is linear in labor; with productivity z_i, firm i produces a quantity $y_i = z_i l_i$ with l_i workers. A firm's productivity z_i is made of two components: an aggregate term common to all firms Ψ, and an idiosyncratic term a_i such that $z_i = \Psi a_i$. The static problem of firm i is therefore to choose y_i, p_i and l_i in every period to maximize instantaneous profits, given its productivity and the inverse demand for its variety. Firms take the equilibrium value of the wage rate, w, as given and solve:

$$\max_{y_i, p_i, l_i} \quad y_i p_i - w l_i \quad \text{subject to} \quad y_i = z_i l_i \quad \text{and} \quad p_i = \left(\frac{Y}{y_i} \right)^{1-\theta} \tag{10}$$

There is a measure L of workers who supply labor inelastically. The equilibrium allocation of labor across monopolists is constrained by the labor market clearing condition: $\int_0^1 l_i \mathrm{d}i = L$. The following lemma summarizes the solution to the static optimization problem of firms.

Lemma 1 (Static equilibrium)**.** *At any instant*

(1) *The optimal output of firm i is* $y_i^* = Y \left(\frac{a_i}{A} \right)^{\frac{1}{1-\theta}}$ *and labor demand is* $l_i^* = \frac{Y}{\Psi} \left(\frac{a_i^\theta}{A} \right)^{\frac{1}{1-\theta}}$.

(2) *Firm i's profits are* $\pi(a_i)^* = Y \left(\frac{a_i}{A} \right)^{\frac{\theta}{1-\theta}} (1-\theta)$ *and its wage bill is* $w l_i^* = Y \left(\frac{a_i}{A} \right)^{\frac{\theta}{1-\theta}} \theta$.

(3) *The wage rate and aggregate output are equal to* $w = \theta A \Psi$ *and* $Y = L A \Psi$*, respectively.*

where $A := \left(\int_0^1 a_i^{\frac{\theta}{1-\theta}} \mathrm{d}i \right)^{\frac{1-\theta}{\theta}}$ *is the idiosyncratic productivity index of the economy.*

Proof. See Appendix G.3 □

[41] $\theta = 1$ means the goods are perfect substitutes, $\theta = 0$ gives a Cobb-Douglas production function and $\theta = -\infty$ means the y_i's perfect complements. Estimates of θ from the literature suggest that its value lies between 0 and 1 *i.e.* intermediate goods are easily substitutable.

Dynamic firm problem. With the static problem of firms solved, I now introduce time subscripts to describe firms' productivity dynamics. Firms' idiosyncratic productivities are stochastic: they follow a geometric Brownian motion with drift rate $\alpha(e_{it}, \beta_{it})$. The drift rate depends on a firm's flow research effort, e_{it} and the type of R&D it performs, described by the indicator β_{it} (for 'basic'). $\beta_{it} = 1$ if it performs basic research and $\beta_{it} = 0$ otherwise. Formally,

$$\frac{\mathrm{d}a_{it}}{a_{it}} = \alpha(e_{it}, \beta_{it})\mathrm{d}t + \nu \mathrm{d}B_t \tag{11}$$

where ν is the standard deviation rate of productivity and $\mathrm{d}B_t$ denotes the standard normal Brownian increment. Mirroring the set up of the estimating equation, the drift rate of firm i's productivity takes the form: $\alpha(e_{it}, \beta_{it}) := e_{it}\phi(\beta_{it})$, where $\phi(\beta_{it})$ is the elasticity of productivity growth to R&D effort. A firm doing basic research ($\beta = 1$) will experience a productivity increase of $e_{it}\phi_1$. On the other hand, if $\beta = 0$ and the firm funds applied research, its productivity increases by $e_{it}\phi_0$. To capture the fact that fundamental R&D does not translate directly into higher productivity and is harder to appropriate by the investing firm, I assume that $\phi_0 > \phi_1$. In other words, firms experience a larger productivity increase when they invest in applied research.

In reality, the 'basicness' of R&D is more a continuum than a clear-cut characteristic. The simple categorization I use here is merely a simplifying assumption. However, modelling the productivity increase from R&D as a function of a continuous measure of R&D 'basicness' can be accommodated by the model.[42]

These productivity dynamics matter to firm owners insofar as they affect their profits. Out of their immediate post-production profits denoted by $\pi(a_{it})^*$, firm owner need to pay taxes at rate τ_t, they need to fund R&D expenses at rate e_{it} and they can consume what remains. They derive log-utility from these post-tax and post-R&D profits so that flow utility is $\ln \pi(a_{it})^*(1 - e_{it} - \tau_t)$.

Finally, the last factor affecting firm owners' utility is the rate of creative destruction. Firm owners can be replaced in two ways. First, they can be replaced by individuals who have found a better version of their variety. In the model, this process of creative destruction materializes through an endogenously determined Poisson rate of exit δ_t. This is the classic Schumpeterian creative destruction and it is an equilibrium quantity. Second, they face a constant and exogenous death rate $\overline{\delta}$ akin the probability of retiring or actually dying. This second mechanism is invariant to the amount of innovation in the economy, unlike δ_t. There is no outside option for firm owners who are replaced.

Putting it all together, a firm owner solves:

[42] For instance, if β_{it} is instead the share of R&D expenditures dedicated to basic research, the results presented in this paper hold if $\phi(\beta_{it})$ is a strictly decreasing function. *I.e.* the more a firm invests in basic research, the less it can generate productivity increments from R&D that it benefits from.

$$
\begin{aligned}
\max_{e_{it},\beta_{it}} \quad & \mathbb{E}_0 \int_0^\infty e^{-\rho t} \ln \pi(a_{it})^*(1-e_{it}-\tau_t)\mathrm{d}t \\
\text{subject to} \quad & \frac{\mathrm{d}a_{it}}{a_{it}} = \alpha(e_{it},\beta_{it})\mathrm{d}t + \nu \mathrm{d}B_t \\
\text{with} \quad & \alpha(e_{it},\beta_{it}) = e_{it}\phi(\beta_{it}) \\
\text{and Poisson rate of exit} \quad & \delta_t + \overline{\delta}
\end{aligned}
\tag{12}
$$

where ρ is the discount rate. Omitting i and t subscripts here as it does not cause confusion, one can write the Hamilton-Jacobi-Bellman equation of a firm with productivity a as

$$
\rho v(a,t) = \max_{e,\beta} \quad \ln \pi(a)^*(1-e-\tau) + \alpha(e,\beta) a v_a(a,t) + \frac{\sigma^2}{2} a^2 v_{aa}(a,t) + v_t(a,t) - (\delta + \overline{\delta}) v(a,t)
\tag{13}
$$

where $v_a(a,t)$ and $v_{aa}(a,t)$ stand for the first and second derivatives of $v(a,t)$ with respect to a, respectively. The value of owning a firm with productivity a is therefore constituted of the utility flow of profits after taxes and R&D expenditures, the change in firm value due to research effort and luck, and the expected loss associated with creative destruction.

6.2. **New ideas.** New ideas play a central role in the model. They are created by researchers hired by firms or by the government and may come from basic or applied research. Beyond the larger impact it has on productivity growth, applied R&D also differs from basic R&D in how it affects ideas. These differences have been documented in the stylized facts of section 3: applied R&D is less likely to generate 'breakthrough' innovations (fact 2) and it is less likely to spill over to the rest of the economy (fact 3). I model these differences explicitly in this section.

Differences between basic and applied R&D. The generation of new ideas depends on the total number of researchers and the type of research they do. When firms spend a share e of their profits on R&D, they hire an aggregate number of researchers $R = e\Pi/w_p$. w_p is the research wage in the private sector, which is different from the wage in the public sector, and $\Pi = \int_0^1 \pi(a_i)^* \mathrm{d}i$ is aggregate profits. If R researchers are doing basic R&D, they get new basic ideas at a Poisson rate of λ ideas per researcher such that $I_1 = \lambda R$. If they conduct applied R&D, they get applied ideas at the same rate: $I_0 = \lambda R$. In other words, generating the same flow of basic or applied ideas is equally hard.

Importantly though, when researchers do basic R&D, a subset of the ideas they generate are breakthroughs, denoted $B_1 \subset I_1$. Breakthroughs from basic R&D arrive at rate λ_1 such that $B_1 = \lambda_1 R$. If instead they work on applied R&D, the breakthrough rate λ_0 is lower and breakthroughs are more rare for the same research effort *i.e.* $B_0 = \lambda_0 R < B_1$. This is consistent with the evidence provided in the stylized facts section that has shown that public R&D (which tends to be more fundamental) produces patents that are more ahead of their time, even after controlling for the cost of research. Table C.14 in the appendix also reports evidence that publicly-funded patents score higher on the popular measure of patent disruptiveness introduced by Kelly *et al.* (2021).

The second key difference between basic and applied R&D is that basic R&D spills over more easily to the rest of the economy. To capture this feature, I assume that λR ideas generated by applied research generate $(\lambda R)^{\varepsilon}$ spillovers to the rest of the economy, while the same number of basic ideas would generate $(\lambda R)^{\gamma}$ spillovers, with $\gamma > \varepsilon$. This captures the feature that an agent will experience the same growth in patents if it invest in basic or applied research (both types of research are equally costly), but when the research is more basic, it spills over more easily to other firms. This is consistent with fact 3 of section 3. The table below summarizes the differences of impact between basic and applied R&D when the same number of researchers (R) works on one or the other.

	Basic		Applied
Researchers		R	
Investment		Rw_p	
Productivity increase	$Rw_p\phi_1/\pi$	$<$	$Rw_p\phi_0/\pi$
Spillovers	$(\lambda R)^{\gamma}$	$>$	$(\lambda R)^{\epsilon}$
Breakthroughs	$\lambda_1 R$	$>$	$\lambda_0 R$

TABLE 5. Impacts of R&D on productivity, spillovers and breakthroughs: Basic v. applied

Spillovers. Applied and basic ideas combine in a Cobb-Douglas aggregator to generate productivity-enhancing spillovers. With R_1 basic researchers and R_0 applied ones, the total amount of spillovers in the economy is given by $\ln(\lambda R_1)^{\gamma}(\lambda R_0)^{\epsilon}$, where the log introduces some curvature in the returns to spillovers. In other words, ideas that can be turned into productivity-enhancing machines or processes are harder to come by when there are already a lot of them.

This functional form captures an important aspect of basic and applied R&D; they are complements in the generation of knowledge spillovers that can be used for productivity growth. For example, the fundamental insights from Shannon's information theory are most useful when combined with the more applied invention of programming languages in order to create the file-compression algorithms that are so crucial to the digital economy. This modelling choice is motivated by several pieces of evidence. First, the SSIV results of section 5 have shown that firm's own R&D, which is more applied, is positively impacted by increases in public R&D spillovers, which tend to be more basic. Second, Moretti *et al.* (2023) have documented that both at the firm and at the industry level, private R&D tends to increase when public R&D increases. Third, evidence from quasi-experimental variation provided by Azoulay *et al.* (2019) and Myers and Lanahan (2022) provide compelling evidence that publicly-funded R&D leads to a large increase in the number of follow-up patents. This aspect of innovative output is consistent with a view of innovation as being both cumulative and combinatorial: discoveries by others make it easier to discover new ideas. The flow of new productivity-enhancing ideas generated through spillovers in the economy at large is then given by $\dot{n}_t := \ln(\lambda R_1)^\gamma(\lambda R_0)^\epsilon$. To simplify the aggregation, spillovers are assumed to be beneficial to all varieties. They are common to all firms and truly capture the wider social benefits that cannot be internalized by firms.

Note that researchers can be in firms, in universities and in governments. They do not necessarily need to *perform* the R&D intramurally *i.e.* where the R&D comes from. This is particularly true for state-funded R&D; A whole 21% of R&D funded by the US federal government was performed by private businesses in 2021, and 28% was performed by universities.[43]

6.3. **Government.** The government also conducts R&D, although with a different objective than firms. It cares about innovation only insofar as it generates breakthroughs findings. Breakthrough innovations are used for whichever cause the government is concerned with at a given instant: like finding a new vaccine to halt the progression of a pandemic, developing new batteries because the price of oil is high, or creating a new weapon.[44] I assume that, at all times, the government needs to satisfy a simple budget constraint that equates expenditures on publicly-funded R&D with aggregate revenue raised by taxing corporate profits. There is no other source of taxation, no government borrowing (no savings technology for that matter) and no other government expenditures. This is a simplification that keeps the model focused and is rather consistent with the recent US fiscal history.[45] In other words, corporate tax totally and exclusively funds government R&D in this model. With its budget raised exclusively from corporate profit tax, the government

[43]Data from the National Science Foundation. Table 6, row 145. Accessed January 10th, 2024. ncses.nsf.gov/data-collections/national-patterns/2021#data

[44]This breakthrough-oriented objective of government-funded research is consistent with US historical evidence. It is best illustrated by the general message of the seminal report 'Science: The Endless Frontier', commissioned by president Franklin D. Roosevelt to translate war-time research efforts into impactful peace-time innovations (Bush, 1945). Its introductory lines read 'Progress in the war against disease depends upon a flow of new scientific knowledge. New products, new industries, and more jobs require continuous additions to knowledge of the laws of nature, and the application of that knowledge to practical purposes. Similarly, our defense against aggression demands new knowledge so that we can develop new and improved weapons. This essential, new knowledge can be obtained only through basic scientific research.'

[45]From the 1980s onward, corporate income tax as a share of US GDP was between 1 and 2.5%, not too far from the 0.7 to 1% of GDP dedicated to publicly-funded R&D.[46] It is slightly less consistent with the immediate postwar period, where corporate income tax revenue accounted for 3.5% of GDP on average between 1950 and 1980, while public R&D was, on average, 1.2% of GDP. Because the two amounts are fairly close, I maintain this simplifying assumption throughout.

then allocates funds to basic and applied research with the aim of maximizing the arrival rate of breakthroughs. Formally, the government's problem is

$$\max_{R_{g1},R_{g0}} \quad \lambda_1 R_{g1} + \lambda_0 R_{g0} \qquad \text{subject to} \qquad \tau\Pi = w_g(R_{g1} + R_{g0}) \tag{14}$$

where R_{g1} and R_{g0} are the numbers of publicly-paid researchers doing basic and applied research, respectively, and w_g is the wage of publicly-paid researchers. In line with the identification assumption of the SSIV exercise, the tax rate τ is taken to be exogenous and is driven by forces outside of the model. A given tax rate fully determines government revenues (and thus public R&D expenditures) given an existing distribution of firm profits.[47]

R&D choices. The different properties of basic and applied R&D, combined with the different objectives of firms and the government lead to a complete specialization of the government in basic research and of the private sector in applied research. Furthermore the R&D effort of firms is constant across the firm size distribution. Proposition 1 below and its proof formalize this result.

Proposition 1 (Endogenous choices of R&D)**.** *Given the problem of firms in* (12) *and the problem of the government in* (14)*:*

(1) $R_g = R_{g1}$*: the government performs basic research, exclusively*
(2) $R_i = R_{i0} \quad \forall i$*: firms perform applied research, exclusively*
(3) *The optimal research effort of firms is unique, independent of firm size and is given by*

$$e^* = 1 - \tau - \frac{1-\theta}{\theta}\frac{\rho + \delta + \bar{\delta}}{\phi_0} \tag{15}$$

Proof. See Appendix G.6 □

The first and second points of this proposition capture the well-known issue of underprovision of public goods. Firms will not be willing to invest in basic R&D if it costs them more, in terms of lost productivity gains, even though it raises aggregate productivity through spillovers by a lot. This prediction of the model is consistent with empirical evidence on corporate science. Arora *et al.* (2021a), for instance, find that firms do little basic research as proxied by their scientific publications; these scientific publications are very rare for firms, even the patent-filing ones.[48]

[47] Using τ as an exogenous variable I can adjust rather than the result of an agent's optimization allows me to make inequality between firms and aggregate productivity growth direct functions of the allocation of R&D resources in the economy. It also makes sense to model it in this way if one is thinking about the government in my model as consisting solely of decision makers in charge of the R&D budgets of federal agencies. These decision makers receive a research budget from another branch of the government who sets τ with a different objective function than theirs.

[48] They find that 2,535 firms out of 4,608 *who already file patents* (55%) have at least a publication in the 1980-2006 period. Moreover, more than 50% of these firms file 0 publications in any given year (table 2, row 6).

Complementing this finding, Akcigit *et al.* (2020) use survey data on the R&D activities of French firms to show that only between 4 and 10% of firms invest in basic research, and only very large firms have non-negligible investments in basic research.[49]

Point (3) of the proposition shows that research effort does not depend on firm size. Because the growth rate of firm's idiosyncradic productivity is constant ($\mathrm{d}a/a = e^{*}\phi_0$), this result yields Gibrat's law, the empirical regularity whereby firms of different sizes grow at the same rate, conditional on survival and age. Moreover, the fact that research effort among R&D-performing firms scales proportionately with firm size finds strong empirical support in the data.[50]

Equation (15) provides intuitive comparative statics. The R&D effort of firms is increasing in the substitubability of varieties θ because productivity gains translate into larger profit gains when θ is high. It also increases in the return to efforts ϕ_0. It decreases in 'impatience' ρ and the probability of being replaced $\delta + \overline{\delta}$ because firm owners enjoy the marginal profit streams over a shorter period of time, in expectation. Finally, and perhaps most importantly for this paper, research effort decreases in the tax rate τ. The negative relationship between research effort and taxes captures the disincentivizing role of taxes on innovation, which has been well documented in the literature. Akcigit *et al.* (2022), for instance, report large elasticities of innovation to the 'keep rate' $(1 - \tau)$ of personal income and corporate taxation in the United States. A 1% increase in the corporate tax keep rate increases patent production by a whole 0.49% according to their estimates.[51]

6.4. **Creative destruction.** Incumbent firm owners can be displaced by workers who discover a better version of their variety. New ideas occur to them through the spillovers of government and private research described earlier such that the Poisson rate of new, viable business ideas at each instant is equal to the amount of spillovers $\dot{n}_t := \ln(\lambda R_1)^{\gamma}(\lambda R_0)^{\epsilon}$. I assume that only a fraction χ of these viable ideas end up being implemented and eventually displace an incumbent. When a worker replaces an incumbent, they inherit the incumbent's idiosyncratic productivity a. The incumbent, once replaced, becomes a worker. This process leaves the productivity distribution of firms unaffected by creative destruction on a BGP: incumbents are immediately replaced by new firm owners with the same productivity. The shape of a productivity distribution under a high equilibrium rate of creative destruction will however be different than under a low one.

The rate of endogenous creative destruction is therefore equal to the rate of spillovers from new ideas, scaled down by the fraction of successfully implemented ideas

[49] Figure 5 of Akcigit *et al.* (2020)

[50] In my sample of firms, investment in R&D typically account for 10% of firm sales and remains a constant share of sales across the firm size distribution.

[51] The corresponding elasticity for the personal income tax rate is even bigger, at 0.8% more patents by 1% increases in the keep rate. Both of these effects, of corporate and personal income tax, are larger at the state level due to migration and R&D re-location responses.

$$\delta := \chi \dot{n}_t \tag{16}$$

More spillovers make the entry of new businesses easier.

Finally, firm owners can also be replaced at an exogenous rate $\overline{\delta}$, already previewed in the firm problem. In that case, they are replaced by new, young firm owners with productivity a_0 set to be equal to the lowest idiosyncratic productivity at a given instant. In other words, a_0 is a reflecting barrier for firm productivity. This exogenous replacement process yields well-behaved productivity distributions (Gabaix, 2009) and is used here for tractability.

6.5. The distribution of firms. At all times, the number of entrants is equal to the number of firms who exit so that the total mass of active firms remains equal to 1. With the creative destruction process described in section 6.4 and the random productivity process (11), the following known result follows;[52] the distribution of firm productivities, $f(a,t)$, evolves over time according to the Kolmogorov Forward Equation (KFE) given by

$$\partial_t f(a,t) = -\overline{\delta} f(a,t) - \alpha \partial_a [a f(a,t)] + \frac{\nu^2}{2} \partial_{aa} [a^2 f(a,t)] \tag{17}$$

where $\partial_t f = \partial f / \partial t$, $\partial_a f = \partial f / \partial a$, and $\partial_{aa} f = \partial^2 f / \partial a^2$. To economize on notation, α stands for $\alpha(e^*, \beta^*)$. On a balanced-growth path, the distribution of firm productivities is stationary *i.e.* $f(a,t) = f(a) \quad \forall a, t$. This stationary distribution must therefore follow the stationary version of the KFE:

$$0 = -\overline{\delta} f(a) - \alpha \partial_a [a f(a)] + \frac{\nu^2}{2} \partial_{aa} [a^2 f(a)] \tag{18}$$

Lemma 2 below shows that the distribution of firm productivities satisfying (18) is a power law. It also shows that the Pareto tail exponent is a function of α (which depends on δ through e).

Lemma 2 (Stationary distribution of firms)**.** *On a balanced-growth path*

- *The stationary distribution of productivities is a power law with density $f(a) = Ca^{-\zeta-1}$ over the support $[a_0, \infty)$.*
- *The Pareto tail exponent ζ is given by*

$$\zeta = -\frac{\alpha}{\nu^2} + \frac{1}{2} + \sqrt{\left(\frac{\alpha}{\nu^2} - \frac{1}{2}\right)^2 + \frac{2\overline{\delta}}{\nu^2}} \tag{19}$$

- *and $C = \zeta {a_0}^{\zeta}$*

Proof. See Appendix G.7 □

[52] See for instance Dixit and Pindyck (1994), p. 89 for a derivation.

ζ is decreasing in α (*i.e.* inequality is increasing in the drift). This means that inequality is accentuated when the rewards to innovating are higher such as when ϕ_0 and θ are higher. Inequality decreases when innovation is disincentivized, for instance when firm owners are more likely to be replaced (higher $\delta + \bar{\delta}$), when the tax rate is higher, or when they are more impatient (higher ρ). The split between public and private R&D will affect inequality through endogenous creative destruction δ: a high probability of being replaced makes firms less likely to grow very large and thus decreases inequality.

Notably, the distribution of a is stationary on a BGP, while the distribution of $\pi(a)$ is a non-stationary travelling wave. This highlights where aggregate growth comes from in the model; spillovers are a 'tide that lift all boats' by multiplicatively scaling up firm idiosyncratic productivities (and thus profits) by Ψ.

6.6. **Equilibrium.** I can now relate aggregate growth and inequality to the allocation of researchers. To do so, I first describe how spillovers affect aggregate growth, I then show how the tax rate determines the key allocation of the model–the split of researchers between public and private R&D–before defining the BGP equilibrium and proving the main result of the paper.

The common productivity term takes the form $\Psi_t = \Gamma^{n_t}$, where Γ is the step size of productivity increments and n_t is the stock of spillovers at time t. This is the standard quality ladder of endogenous growth models. Hence firm productivity is $z_{it} = \Gamma^{n_t} a_{it}$. From lemma 1, the aggregate productivity growth rate of the economy is the same as that of GDP per capita and is equal to

$$g = \dot{n}_t \ln \Gamma \tag{20}$$

where $\dot{n}_t = \ln(\lambda R_1)^{\gamma}(\lambda R_0)^{\varepsilon}$ as established earlier. Taking logs and time differences of $z_{it} = \Gamma^{n_t} a_{it}$, I get the estimating equation of section 4.

$$\Delta \ln(z_{it}) = \phi_0 \underbrace{e_{it}}_{\substack{\text{own} \\ \text{R\&D flow}}} + \gamma \underbrace{\ln(\lambda R_1)}_{\substack{\text{flow of} \\ \text{basic ideas}}} + \varepsilon \underbrace{\ln(\lambda R_0)}_{\substack{\text{flow of} \\ \text{applied ideas}}} \tag{21}$$

Researchers hired by firms receive a proportional wage premium Λ over what they would earn if they were funded by the government, such that $w_p = \Lambda w_g$. This is a reduced-form way of capturing a well-documented feature of the labor market: private-sector workers typically enjoy a 5-to-30% wage premium over what they would earn in the public sector (Murphy *et al.*, 2020). The research wage bill of firms is $e\Pi = w_p R_p$ and the research wage bill of the government is $\tau\Pi = w_g R_p$. Given an exogenous tax rate τ and the research labor constraint $R = R_g + R_p$, the wage rates for researchers adjusts to clear the market. The number of researchers in each sector is then given by two simple relationships;

$$R_g = \frac{R}{e/\Lambda\tau + 1} \qquad \text{and} \qquad R_p = \frac{R}{\Lambda\tau/e + 1} \tag{22}$$

The comparative statics are as follows. Publicly-funded researchers become more numerous when τ increases. They also become more numerous when the premium paid to private researchers is bigger, all else equal, because firms can hire fewer researchers and thus leave more of them to the public sector. In contrast, a bigger research effort by firms increases the number of private researchers to the detriment of publicly-funded ones.

The BGP equilibrium is characterized by 12 key endogenous variables–$Y, y_i, a_i, L, l_i, e, R_p, R_g, \dot{n}, \delta, \beta_g, \beta_i$– and an equal number of equations, listed in Table G.23 in the appendix. The definition of a BGP equilibrium is standard. Given a tax rate τ, (i) firm owners choose y_i, l_i, e_i and β_i to maximize the present discounted value of owning a firm, (ii) the government chooses the type of R&D that maximizes the arrival rate of breakthroughs, (iii) workers and researchers supply labour inelastically and (iv) the wage rates of workers and researchers clear their respective labor markets. These interactions yield two coupled functions $(f, v) : [a_0, \infty) \to \mathbb{R}$ which are the stationary density of firm productivities and the value function of firm owners. On a BGP, aggregate productivity, wages and output per capita grow at g. Incumbents' profits and wage bills grow at $g + \frac{\theta}{1-\theta} e\phi_0$, on average, as long as they do not exit.

Through its effect on the allocation of researchers to basic (public) R&D and applied (private) R&D, τ affects the strength of spillovers in the economy, which in turn affects aggregate growth via Γ^{n_t} and inequality via δ. Proposition 2 below shows how growth and firm inequality evolve as a function of the allocation of researchers to basic and applied research.

Proposition 2 (Taxes, growth and inequality)**.** *On balanced-growth paths:*

(1) *Inequality of productivity between firms is U-shaped in the share of researchers in the private sector.*

(2) *The aggregate productivity growth rate of the economy is inverted U-shaped in the share of researchers in the private sector.*

(3) *There is a unique, growth-maximizing and inequality-minimizing tax rate given by*

$$\tau^* = \frac{\gamma e^*}{\varepsilon\Lambda} \tag{23}$$

and the associated share of government researchers is $\frac{{R_g}^*}{R} = \frac{1}{\varepsilon/\gamma + 1}$

Proof. See Appendix G.8. □

Two properties of spillovers are key to explaining proposition 2: the complementarity between the two types of R&D and the decreasing marginal impact of each on the flow of overall spillovers.

At low levels of tax, spillovers are dominated by spillovers from private research because the government has little resources to fund basic research and because R&D by firms is strongly incentivized by low taxes. As the tax rate rises, the level of spillovers increases because public spillovers get larger and have a high marginal impact on overall spillovers. At τ^*, the marginal impacts of basic and applied spillovers are equalized. Finally, when the tax rate is getting too high, research by private firms is disincentivized and private spillovers fall out of balance. Aggregate spillovers are falling too.

The growth-maximizing tax rate τ is increasing in the strength of publicly-funded spillovers (γ) and decreasing in the strength of privately-funded spillovers (ε). Interestingly, it is increasing in private research effort: just like private R&D is complementary to public R&D, the reverse is also true and high levels of private R&D make public R&D more impactful. Finally, it decreases in the private wage premium because a lower tax rate is needed to fund the optimal number of public researchers when Λ is low.

6.7. **Calibration.** I now evaluate the ability of the model to explain (part of) the decline in TFP and the increase in firm inequality, from 1950 to 2017. To do so, I calibrate it with standard parameter values taken from the literature such that it matches the growth rate of TFP (g) and the Pareto tail exponent (ζ) in the immediate postwar period. The model is stylized and the causes of the secular decline in productivity in the US are multiple. My goal is therefore not to explain all of the TFP deceleration in the US postwar history but to highlight the role public R&D may play as one cause of the slowdown. Complementary explanations of the decline in TFP growth and the rise in firm inequality are discussed at the end of this section. I present here a sequence of BGP equilibria and I elaborate more on the calibration exercise in Appendix H.

Set up. The tractability of the model makes the calibration exercise transparent. I have indeed obtained closed-form expressions for the two quantities I am interested in; the Pareto tail exponent of inequality between firms (19) and the growth rate of aggregate productivity (20). Given parameter values of ν, θ, ϕ_0, γ, ε, ρ, Γ, λ, Λ, χ, $\overline{\delta}$ and a time series of tax rates τ_t, I can obtain the values of equilibrium quantities e^*, δ, $\dot{n}_t$, which give me a sequence of values for g and ζ.

Values of ρ, ν, θ, Λ, Γ and $\overline{\delta}$ are taken from the macro literature, γ and ε are taken from my empirical exercises, χ is calibrated so that the exit rate takes on a realistic value, λ and ϕ_0 are internally calibrated to match the values of g and ζ at the beginning of the period. τ, the main exogenous input to the model is set equal to the effective corporate tax rate in the US at the beginning of the period. It is then set to follow the share of public R&D in overall R&D. The tax rate set in this way closely follow the historical time series of the effective corporate tax rate (see Figure 20 in the

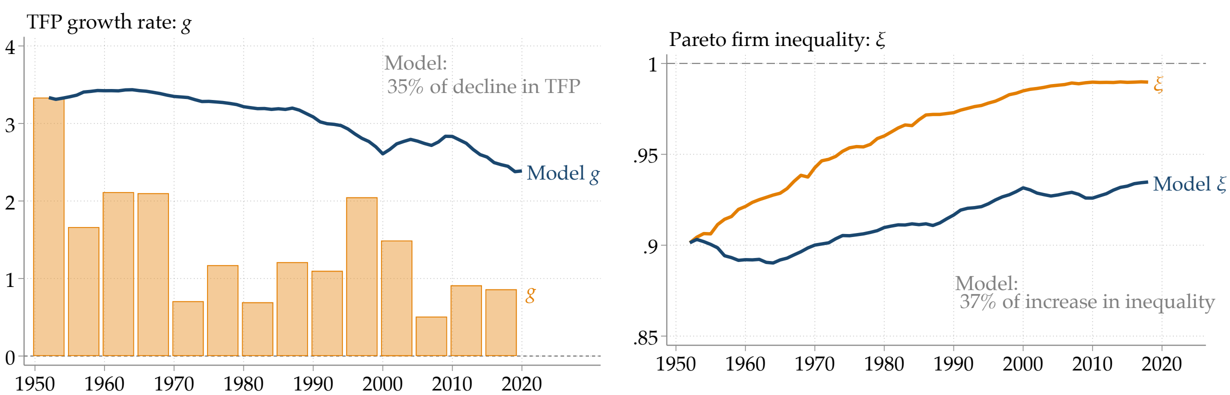
Appendix). Appendix H describes the data sources used in the exercise and provides more information about the calibration procedure. Table 6 lists the parameter values and motivates their choices.

Parameter	Value	Source/Meaning
Government		
τ	0.34	Set equal to the effective corporate tax rate in 1947
		Then inferred from the changes in the public R&D budget share of total R&D in the US
Λ	1.25	Public-private wage gap at 50th percentile from Murphy *et al.* (2020), p. 284
Firms		
ν	0.4	Luttmer (2007), p.1132
ϕ_0	0.1	Middle-of-the-road value of estimates of VA elasticity to R&D, from review by Hall *et al.* (2010)
ρ	0.01	Standard
$\overline{\delta}$	0.035	Employment-weighted exit rate from Decker *et al.* (2016) (p. 9)
ζ_0	1.109	Observed in the data (tail exponent in 1952)
g_0	0.033	Observed in the data (average TFP growth rate in 1950-1955)
Γ	1.4	Jones and Kim (2018), p.1809
θ	3/4	Standard
Research and spillovers		
γ	0.04	Table 1, column (4)
ε	$\gamma/3$	A third of γ, from section 5.2
λ	0.12	Internally calibrated to match ζ_0
χ	0.05	Internally calibrated to match g_0

TABLE 6. Calibrated parameter values

Results. The results of the calibration exercise suggest that the decline in the share of GDP dedicated to public R&D can explain a substantial share of the deceleration in TFP and a substantial share of the rise in inequality between firms. Starting with TFP growth, Figure 7a shows how the growth rate of aggregate TFP predicted by the model compares to the data. Both series start at the same growth rate of 3.3% in the early 1950s, by construction. Immediately after, the growth rate predicted by the model increases as spillovers from the rise in public R&D in the 1950s bear fruits and drive private firms' productivity up. Soon after though, the balance of spillovers starts to tilt toward spillovers from private R&D. Because the elasticity of applied spillovers (from the private sector) is lower than that of basic spillovers (from the public sector), the growth-maximizing mix of spillovers will have more public than private R&D. The model reflects this shift by lowering the equilibrium growth rate of TFP from the 70s to present days. Over the entire period, g_{model} decreases from 3.33% to 2.46%, a 0.86 percentage point decrease. In the data, TFP growth fell from 3.33% to 0.86% (-2.47pp). In other words, the model accounts for slightly more than a third of the fall in TFP growth over the period (35%).

Turning to inequality between firms, the historical data shows a continuous increase in inequality from 1952 to 2018, as documented by Kwon *et al.* (2022) and shown in 7b. It is more intuitive to refer to power law inequality, defined as $\xi := 1/\zeta$, when describing changes in inequality between firms rather than to the Pareto tail exponent ζ. Higher levels of inequality yield higher ξ and the calibration exercise uses power law inequality rather than the Pareto tail exponent as an object of interest. I rely on Kwon *et al.* (2022)'s series on corporate assets here as this series spans the entire period I am interested in. Series on receipts and net income (which would have a more direct counterpart in my model) are unfortunately not available for the full period. It is however notable that all three series on inequality of assets, receipts and net income yield almost identical Pareto exponents over the periods when they overlap. The increase in inequality predicted by the model, in contrast to the data, is not monotonic. After starting from the same level in the beginning of the 1950s (by construction), it decreases down to its lowest level in the middle of the 1960s. The model ascribes this decrease in inequality to the rise of spillovers in the late 1950s and early 1960s. After this temporary fall, inequality increases until 2017 up to a value of ξ implying that the top 1% share of firms by assets owns 72% of all firm assets. The corresponding figure in the data is 95% in 2018. In sum, the model can explain 37% of the rise in inequality between firms from the 1950s to 2017.



(A) TFP growth

(B) Firm inequality $\xi = \zeta^{-1}$

FIGURE 7. Calibration results

Notes: Parameter values are either estimated in my empirical exercises or taken from the literature. See Table 6 for more details. The Pareto firm inequality parameter ξ can be given an intuitive interpretation by using the following property of Pareto distributions. The top share of the p% biggest firms is given by $(100/p)^{\xi-1}$. Applying this insight to the empirical time series of 7b, one gets that the top 1% share of firm assets was around 60% in the early 1950s and 95% in the late 2010s.

Discussion. While the calibration exercise suggests that the change in R&D funding can account for a large part of the decline in TFP growth, it is unlikely to be the only driver of long-term changes in TFP. An alternative, yet related, explanation builds upon the idea that 'ideas are getting harder to find' (Bloom *et al.*, 2020): innovation-driven improvements in TFP were easier to achieve in previous decades. My theory offers a potential cause of the 'ideas are harder to find' hypothesis: maybe the rate of growth of ideas is a function of the type of research conducted by a society, applied or basic. The steady decline in public R&D in the US could be a cause of the fact that ideas are harder to find.

Over shorter time horizons, other theories may be better at explaining variations in TFP. TFP growth is indeed fairly cyclical and the decline in public R&D is more of a long-term trend. De Ridder (Forthcoming), for instance, ascribes the large productivity growth of the late 1990s and its subsequent decline to the rise of corporate investments in intangible assets (like software). Alternatively, Liu *et al.* (2022) build a theory linking the decline in interest rates to a stronger investment response by market leaders than by followers, which leads to a joint rise in concentration and a slowdown of growth.

Alongside these theories, my model and its calibration serve as a proof of concept that the decline in public R&D may be an alternative (and complementary) mechanism behind the fall in productivity growth and the rise in firm inequality.

7. Conclusion

Through the lens of a 70-year panel of firms matched to patents, two quasi-experimental IV strategies and a calibrated model of growth, this project has provided evidence that the split between publicly and privately-funded R&D matters for the intensity of knowledge spillovers in an economy. It has also shown that this public v. private partition has an impact on the growth rate of productivity and on how unequal the firm size distribution is. The core distinction between publicly and privately-funded R&D that drives these results stems from the fact that the former is more *fundamental* than the latter. The two empirical exercises show that public R&D positively impacts private firms' productivity growth through spillovers over the long run (SSIV), and there is tentative evidence that this impact is at least twice as big as that of private R&D (patent examiner instrument). This difference of impact matters in the aggregate, as evidenced by the fact that the decline in public R&D in the US can explain a third of the deceleration in TFP from the 1950s to present days, and a third of the rise in inequality between firms, according to my calibrated model of growth. While the causes of the secular decline in TFP growth are multifaceted, my findings point to an underappreciated factor: public R&D as a source of impactful spillovers for private firms.

This line of research contributes to the ongoing debate in the US and Europe about the role of public R&D investments in fostering productivity growth and the relevance of basic R&D investments in industrial policy. However, the extent to which the conclusions of this project can be generalized to countries other than the US (or other advanced economies) is an open question. The American economy over the post-WWII period is indeed unique in two important ways. First, the US has been at the technological frontier, or near it, in many domains over this period. In this respect, fundamental R&D funded by the government may be the most appropriate tool to push the frontier. For instance, Ahmadpoor and Jones (2017) and the stylized facts of section 3 provide evidence that patents drawing heavily on scientific papers tend to be the most impactful (as measured by their citation counts). In contrast, funding or subsidizing applied R&D may be the most adequate strategy for an economy trying to catch up with the frontier. Second, the US innovation system has been distinctively capable of translating insights from basic R&D into innovative products and services due to a strong innovation pipeline from universities to corporate labs and to final production, at least until the 1980s (Arora *et al.*, 2020).

Understanding the roles government can play in accelerating productivity growth is a fertile ground for future research. In particular, the research presented here can be extended in several ways. Valuable extension of this work include a deeper exploration of the specific mechanisms whereby publicly-funded R&D generates more spillovers. Previous evidence suggests that the different incentives researchers face when their work is funded by public versus private money may be important (Babina *et al.*, 2023). The exact ways in which these spillover operate (through the movement of scientists or public-private partnerships for instance) is another question worthy of exploration. Finally, it would also be interesting to jointly assess the respective impacts of publicly-funded R&D spillovers and government demand shocks on productivity growth, within a unified empirical framework.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451. [*Cited on page 29.*]

Acs, Z. J., D. B. Audretsch, and M. P. Feldman (1994): "R&D Spillovers and Recipient Firm Size," *The Review of Economics and Statistics*, 76, 336–340. [*Cited on pages 2, 15, and 34.*]

Adão, R., M. Kolesár, and E. Morales (2019): "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 134, 1949–2010. [*Cited on pages 2, 20, 26, 28, 30, 32, 33, 36, and 92.*]

Adams, J. D. (1990): "Fundamental stocks of knowledge and productivity growth," *Journal of political economy*, 98, 673–702. [*Cited on page 18.*]

Agarwal, R. and P. Gaule (2022): "What drives innovation? Lessons from COVID-19 R&D," *Journal of Health Economics*, 82, 102591. [*Cited on page 14.*]

Agrawal, A., C. Rosell, and T. S. Simcoe (2014): *Do Tax Credits Affect R & D Expenditures by Small Firms?: Evidence from Canada*. [*Cited on page 68.*]

Ahmadpoor, M. and B. F. Jones (2017): "The dual frontier: Patented inventions and prior scientific advance," *Science*, 357, 583–587. [*Cited on pages 18, 54, and 90.*]

Akcigit, U. and S. T. Ates (2019): "Ten facts on declining business dynamism and lessons from endogenous growth theory," Tech. rep., National Bureau of Economic Research. [*Cited on page 4.*]

——— (2022): "What Happened to US Business Dynamism?" Tech. rep. [*Cited on page 33.*]

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva (2022): "Taxation and Innovation in the Twentieth Century," *The Quarterly Journal of Economics*, 137, 329–385. [*Cited on page 46.*]

Akcigit, U., D. Hanley, and N. Serrano-Velarde (2020): "Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth," *The Review of Economic Studies*, 88, 1–43. [*Cited on pages 4, 46, and 90.*]

Akcigit, U. and W. R. Kerr (2018): "Growth through heterogeneous innovations," *Journal of Political Economy*, 126, 1374–1443. [*Cited on page 5.*]

Albert, M. B., D. Avery, F. Narin, and P. McAllister (1991): "Direct validation of citation counts as indicators of industrially important patents," *Research policy*, 20, 251–259. [*Cited on page 80.*]

Antolin-Diaz, J. and P. Surico (2022): *The long-run effects of government spending*, Centre for Economic Policy Research. [*Cited on pages 4 and 20.*]

Arora, A., S. Belenzon, and A. Patacconi (2018): "The decline of science in corporate R&D," *Strategic Management Journal*, 39, 3–32. [*Cited on page 13.*]

Arora, A., S. Belenzon, A. Patacconi, and J. Suh (2020): "The Changing Structure of American Innovation: Some Cautionary Remarks for Economic Growth," *Innovation Policy and the Economy*, 20, 39–93. [*Cited on page 54.*]

Arora, A., S. Belenzon, and L. Sheer (2021a): "Knowledge spillovers and corporate investment in scientific research," *American Economic Review*, 111, 871–98. [*Cited on pages 5, 45, and 90.*]

——— (2021b): "Matching patents to compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures," *Research Policy*, 50, 104217. [*Cited on pages 5, 6, 7, 75, 76, and 77.*]

Audretsch, D. B. and M. Vivarelli (1996): "Firms size and R&D spillovers: Evidence from Italy," *Small Business Economics*, 8, 249–258. [*Cited on page 34.*]

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 135, 645–709. [*Cited on page 33.*]

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2019): "Public R&D investments and private-sector patenting: evidence from NIH funding rules," *The Review of economic studies*, 86, 117–152. [*Cited on pages 3, 13, 27,* *and 44.*]

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, Entrepreneurship, and Publications*,” *The Quarterly Journal of Economics*, 138, 895–954. [*Cited on pages 12, 14, and 54.*]

BALCONI, M., S. BRUSONI, AND L. ORSENIGO (2010): “In defence of the linear model: An essay,” *Research policy*, 39, 1–13. [*Cited on page 90.*]

BELENZON, S. AND L. C. CIOACA (2022): “Guaranteed Public Demand and Corporate Scientific Research,” . [*Cited on page 4.*]

BELENZON, S. AND M. SCHANKERMAN (2013): “Spreading the Word: Geography, Policy, and Knowledge Spillovers,” *The Review of Economics and Statistics*, 95, 884–903. [*Cited on page 3.*]

BENSON, C. L. AND C. L. MAGEE (2015): “Quantitative determination of technological improvement from patent data,” *PloS one*, 10, e0121635. [*Cited on page 80.*]

BERGEAUD, A., G. CETTE, AND R. LECAT (2016): “Productivity trends in advanced countries between 1890 and 2012,” *Review of Income and Wealth*, 62, 420–444. [*Cited on pages 64 and 106.*]

BERGEAUD, A., A. GUILLOUZOUIC, E. HENRY, AND C. MALGOUYRES (2022a): “From public labs to private firms: magnitude and channels of R&D Spillovers,” . [*Cited on page 3.*]

BERGEAUD, A., C. VERLUISE, *et al.* (2022b): “A new Dataset to Study a Century of Innovation in Europe and in the US,” Tech. rep., Centre for Economic Performance, LSE. [*Cited on page 6.*]

BESSEN, J. (2009): “NBER PDP project user documentation,” *Data available at: https://sites. google. com/site/patentdataproject/Home/downloads*. [*Cited on page 77.*]

BLOOM, N., C. I. JONES, J. VAN REENEN, AND M. WEBB (2020): “Are ideas getting harder to find?” *American Economic Review*, 110, 1104–1144. [*Cited on pages 16 and 53.*]

BLOOM, N., M. SCHANKERMAN, AND J. VAN REENEN (2013): “Identifying technology spillovers and product market rivalry,” *Econometrica*, 81, 1347–1393. [*Cited on pages 3, 6, 16, and 17.*]

BORUSYAK, K., P. HULL, AND X. JARAVEL (2022): “Quasi-experimental shift-share research designs,” *The Review of Economic Studies*, 89, 181–213. [*Cited on pages 2 and 23.*]

BRUNT, L., J. LERNER, AND T. NICHOLAS (2012): “Inducement Prizes and Innovation,” *The Journal of Industrial Economics*, 60, 657–696. [*Cited on page 14.*]

BUERA, F. J. AND R. E. LUCAS (2018): “Idea flows and economic growth,” *Annual Review of Economics*, 10, 315–345. [*Cited on page 4.*]

BUSH, V. (1945): “Science: the endless frontier, a report to the President, US Government Printing Office,” Tech. rep., Washington, DC. [*Cited on pages 44 and 90.*]

CHEN, Z. (2022): “Economic Growth and the Rise of Large Firms,” . [*Cited on page 106.*]

CHEN, Z., Z. LIU, J. C. SUÁREZ SERRATO, AND D. Y. XU (2021): “Notching R&D Investment with Corporate Income Tax Cuts in China,” *American Economic Review*, 111, 2065–2100. [*Cited on page 69.*]

CRESCENZI, R., A. DYÈVRE, AND F. NEFFKE (2022): “Innovation Catalysts: How Multinationals Reshape the Global Geography of Innovation,” *Economic Geography*, 0, 1–29. [*Cited on page 80.*]

CZARNITZKI, D., K. HUSSINGER, AND C. SCHNEIDER (2011): ““Wacky” patents meet economic indicators,” *Economics Letters*, 113, 131–134. [*Cited on page 80.*]

DE RASSENFOSSE, G., A. JAFFE, AND E. RAITERI (2019): “The procurement of innovation by the US government,” *PloS one*, 14, e0218927. [*Cited on page 27.*]

De Ridder, M. (Forthcoming): "Market power and innovation in the intangible economy," *American Economic Review*. [*Cited on page 53.*]

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016): "Where has all the skewness gone? The decline in high-growth (young) firms in the U.S." *European Economic Review*, 86, 4–23, the Economics of Entrepreneurship. [*Cited on page 51.*]

Department of Veterans Affairs (2022): "Transforming Strategy into Action: FY 22 Strategic Plan," Tech. rep. [*Cited on page 8.*]

Dinlersoz, E., S. Kalemli-Ozcan, H. Hyatt, and V. Penciakova (2018): "Leverage over the life cycle and implications for firm growth and shock responsiveness," Tech. rep., National Bureau of Economic Research. [*Cited on page 6.*]

Dixit, A. K. and R. S. Pindyck (1994): *Investment under uncertainty*, Princeton university press. [*Cited on page 47.*]

Dyèvre, A. and O. Seager (forthcoming): "Matching Patents to Publicly Listed Firms in the US: 1950-2020," Tech. rep., LSE Working paper. [*Cited on pages 5, 6, 7, and 75.*]

Farre-Mensa, J., D. Hegde, and A. Ljungqvist (2020): "What is a patent worth? Evidence from the US patent "lottery"," *The Journal of Finance*, 75, 639–682. [*Cited on page 24.*]

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups, and US Welfare," *Journal of Political Economy*, 125, 1040–1074. [*Cited on page 34.*]

Feng, J. and X. Jaravel (2020): "Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation," *American Economic Journal: Applied Economics*, 12, 140–81. [*Cited on pages 2 and 24.*]

Fieldhouse, A. and K. Mertens (2023): "The Returns to Government R&D: Evidence from US Appropriations Shocks," . [*Cited on pages 4, 5, 22, 29, 31, 92, 93, and 94.*]

Fleming, L., H. Greene, G. Li, M. Marx, and D. Yao (2019): "Government-funded research increasingly fuels innovation," *Science*, 364, 1139–1141. [*Cited on pages 6, 7, 75, and 78.*]

Furman, J. (2015): "Business investment in the united states: Facts, explanations, puzzles, and policies," *Reviving Private Investment" Remarks at the Progressive Policy Institute Washington, DC*. [*Cited on page 33.*]

Gabaix, X. (2009): "Power laws in economics and finance," *Annu. Rev. Econ.*, 1, 255–294. [*Cited on page 47.*]

Gaule, P. (2018): "Patents and the success of venture-capital backed startups: Using examiner assignment to estimate causal effects," *The Journal of Industrial Economics*, 66, 350–376. [*Cited on page 2.*]

Godin, B. (2006): "The Linear Model of Innovation: The Historical Construction of an Analytical Framework," *Science, Technology, & Human Values*, 31, 639–667. [*Cited on page 90.*]

Graham, S. J., A. C. Marco, and R. Miller (2018): "The USPTO patent examination research dataset: A window on patent processing," *Journal of Economics & Management Strategy*, 27, 554–578. [*Cited on page 79.*]

Griliches, Z. (1979): "Issues in assessing the contribution of research and development to productivity growth," *The bell journal of economics*, 92–116. [*Cited on page 15.*]

——— (1992): "The Search for R&D Spillovers," *The Scandinavian Journal of Economics*, 94, S29–S47. [*Cited on page 3.*]

Gross, D. P. and B. N. Sampat (2024): "The Government Patent Register: A New Resource for Measuring US Government-Funded Patenting," Tech. rep., National Bureau of Economic Research. [*Cited on page 7.*]

Gruber, J. and S. Johnson (2019): *Jump-starting America: How breakthrough science can revive economic growth and the American dream*, Hachette UK. [*Cited on page 4.*]

Guenther, G. (2022): "Federal Research Tax Credit: Current Law and Policy Issues," . [*Cited on page 68.*]

Gutiérrez, G. and T. Philippon (2017): "Declining Competition and Investment in the US," Tech. rep., National Bureau of Economic Research. [*Cited on page 33.*]

HALL, B. H., A. JAFFE, AND M. TRAJTENBERG (2005): "Market Value and Patent Citations," *The RAND Journal of Economics*, 36, 16–38. [*Cited on page 80.*]

HALL, B. H., A. B. JAFFE, AND M. TRAJTENBERG (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," Working Paper 8498, National Bureau of Economic Research. [*Cited on page 75.*]

HALL, B. H., J. MAIRESSE, AND P. MOHNEN (2010): "Chapter 24 - Measuring the Returns to R&D," in *Handbook of the Economics of Innovation, Volume 2*, ed. by B. H. Hall and N. Rosenberg, North-Holland, vol. 2 of *Handbook of the Economics of Innovation*, 1033–1082. [*Cited on page 51.*]

HASSLER, J., P. KRUSELL, AND C. OLOVSSON (2021): "Directed Technical Change as a Response to Natural Resource Scarcity," *Journal of Political Economy*, 129, 3039–3072. [*Cited on page 14.*]

HSIEH, C.-T. AND E. ROSSI-HANSBERG (2023): "The Industrial Revolution in Services," *Journal of Political Economy Macroeconomics*, 1, 3–42. [*Cited on page 33.*]

IARIA, A., C. SCHWARZ, AND F. WALDINGER (2018): "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science*," *The Quarterly Journal of Economics*, 133, 927–991. [*Cited on page 3.*]

ILZETZKI, E. (2022): "Learning by necessity: Government demand, capacity constraints, and productivity growth," *LSE, March*. [*Cited on pages 4, 14, and 20.*]

JAFFE, A. B. (1986): "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *The American Economic Review*, 76, 984–1001. [*Cited on page 16.*]

——— (1989): "Real Effects of Academic Research," *The American Economic Review*, 79, 957–970. [*Cited on page 3.*]

JAFFE, A. B. AND G. DE RASSENFOSSE (2017): "Patent citation data in social science research: Overview and best practices," *Journal of the Association for Information Science and Technology*, 68, 1360–1374. [*Cited on page 80.*]

JAFFE, A. B. AND M. TRAJTENBERG (1999): "International Knowledge Flows: Evidence From Patent Citations," *Economics of Innovation and New Technology*, 8, 105–136. [*Cited on page 65.*]

JAFFE, A. B., M. TRAJTENBERG, AND M. S. FOGARTY (2000): "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *The American Economic Review*, 90, 215–218. [*Cited on page 9.*]

JANEWAY, W. H. (2018): *Doing capitalism in the innovation economy: Reconfiguring the three-player game between markets, speculators and the state*, Cambridge University Press. [*Cited on page 4.*]

JIN, C., Y. MA, AND B. UZZI (2021): "Scientific prizes and the extraordinary growth of scientific topics," *Nature communications*, 12, 5619. [*Cited on page 14.*]

JONES, C. I. (1995): "R&D-based models of economic growth," *Journal of political Economy*, 103, 759–784. [*Cited on page 4.*]

——— (2022): "The past and future of economic growth: A semi-endogenous perspective," *Annual Review of Economics*, 14, 125–152. [*Cited on page 4.*]

JONES, C. I. AND J. KIM (2018): "A Schumpeterian model of top income inequality," *Journal of political Economy*, 126, 1785–1826. [*Cited on pages 3, 39, 51, and 102.*]

JONES, C. I. AND J. C. WILLIAMS (1998): "Measuring the Social Return to R&D," *The Quarterly Journal of Economics*, 113, 1119–1135. [*Cited on page 4.*]

KANTOR, S. AND A. WHALLEY (2022): "Moonshot: Public R&D and Growth," *Working paper*. [*Cited on pages 4 and 5.*]

KELLY, B., D. PAPANIKOLAOU, A. SERU, AND M. TADDY (2021): "Measuring technological innovation over the long run," *American Economic Review: Insights*, 3, 303–20. [*Cited on pages 11, 43, 79, and 85.*]

KLING, J. R. (2006): "Incarceration length, employment, and earnings," *American Economic Review*, 96, 863–876. [*Cited on page 2.*]

KOGAN, L., D. PAPANIKOLAOU, A. SERU, AND N. STOFFMAN (2017): "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132, 665–712. [*Cited on pages 7, 75, and 78.*]

KWON, S. Y., Y. MA, AND K. ZIMMERMANN (2022): "100 Years of Rising Corporate Concentration," *Working paper*. [*Cited on pages 2, 33, 52, and 106.*]

LAFOND, F. AND D. KIM (2019): "Long-run dynamics of the US patent classification system," *Journal of Evolutionary Economics*, 29, 631–664. [*Cited on page 82.*]

LEMLEY, M. A. AND B. SAMPAT (2012): "Examiner characteristics and patent office outcomes," *Review of economics and statistics*, 94, 817–827. [*Cited on page 2.*]

LEV, B. AND G. MANDELKER (1972): "The microeconomic consequences of corporate mergers," *Journal of Business*, 85–104. [*Cited on pages 6, 76, and 77.*]

LIU, E. AND S. MA (2023): "Innovation Networks and R&D Allocation," *Working Paper*. [*Cited on page 65.*]

LIU, E., A. MIAN, AND A. SUFI (2022): "Low interest rates, market power, and productivity growth," *Econometrica*, 90, 193–221. [*Cited on page 53.*]

LUCAS, R. E. (2009): "Ideas and growth," *Economica*, 76, 1–19. [*Cited on page 4.*]

LUCKING, B., N. BLOOM, AND J. VAN REENEN (2019): "Have R&D Spillovers Declined in the 21st Century?" *Fiscal Studies*, 40, 561–590. [*Cited on pages 3 and 5.*]

LUTTMER, E. G. (2007): "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 122, 1103–1144. [*Cited on pages 3, 39, and 51.*]

MACLAURIN, W. R. (1953): "The sequence from invention to innovation and its relation to economic growth," *The Quarterly Journal of Economics*, 67, 97–111. [*Cited on page 90.*]

MANSFIELD, E. (1991): "Academic research and industrial innovation," *Research policy*, 20, 1–12. [*Cited on page 17.*]

MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60, 531–542. [*Cited on page 19.*]

MARX, M. AND A. FUEGI (2020): "Reliance on science: Worldwide front-page patent citations to scientific articles," *Strategic Management Journal*, 41, 1572–1594. [*Cited on page 18.*]

——— (2022): "Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations," *Journal of Economics & Management Strategy*, 31, 369–392. [*Cited on page 9.*]

MATCHAM, W. AND M. SCHANKERMAN (2023): "Screening Property Rights for Innovation," . [*Cited on page 10.*]

MAZZUCATO, M. (2015): *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*, Hachette UK. [*Cited on page 4.*]

——— (2021): *Mission economy: A moonshot guide to changing capitalism*, Penguin UK. [*Cited on page 14.*]

MELASS, A., C. KESTEVEN, AND M. HORTON (2021): "Sustaining California's Innovation Economy through Investment in R&D," . [*Cited on page 68.*]

MEZZANOTTI, F. AND T. SIMCOE (2023): "Innovation and Appropriability: Revisiting the Role of Intellectual Property," Tech. rep. [*Cited on pages 6 and 35.*]

MILLER, R. (2020): "Technical documentation for the 2019 patent examination research dataset (PatEx) release," . [*Cited on page 8.*]

MONATH, N., C. JONES, AND S. MADHAVAN (2021): "Disambiguating Inventors, Assignees, and Locations," *American Institute for Research working paper*. [*Cited on page 78.*]

MORETTI, E., C. STEINWENDER, AND J. VAN REENEN (2019): "The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers," Working Paper 26483, National Bureau of Economic Research. [*Cited on pages 5 and 29.*]

——— (2023): "The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers," *The Review of Economics and Statistics*, 1–46. [*Cited on pages 4 and 43.*]

MOSER, P., A. VOENA, AND F. WALDINGER (2014): "German Jewish Émigrés and US Invention," *American Economic Review*, 104, 3222–55. [*Cited on page 3.*]

MOWERY, D. C., R. R. NELSON, B. N. SAMPAT, AND A. A. ZIEDONIS (2015): *Ivory tower and industrial innovation: University-industry technology transfer before and after the Bayh-Dole Act*, Stanford University Press. [*Cited on page 18.*]

MURPHY, P., D. BLACKABY, N. O'LEARY, AND A. STANEVA (2020): "Understanding What Has Been Happening to the Public-Sector Pay Premium in Great Britain: A Distributional Approach Based on the Labour Force Survey," *British Journal of Industrial Relations*, 58, 273–300. [*Cited on pages 48 and 51.*]

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, 2393–2423. [*Cited on pages 3 and 44.*]

NARIN, F. (1995): "Patents as indicators for the evaluation of industrial research output," *Scientometrics*, 34, 489–496. [*Cited on page 80.*]

NATIONAL SCIENCE BOARD, N. (1975): *Science Indicators, 1974: Report of the National Science Board*, US Government Printing Office. [*Cited on page 18.*]

NELSON, R. R. (1959): "The simple economics of basic scientific research," *Journal of political economy*, 67, 297–306. [*Cited on page 90.*]

NSF NATIONAL SCIENCE BOARD (2022): "Research and Development: U.S. Trends and International Comparisons," Science and Engineering Indicators NSB-2022-5, National Science Foundation, Alexandria, VA. [*Cited on page 67.*]

OECD (2021): "R&D Tax Incentives: United States, 2021," . [*Cited on pages 66 and 67.*]

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297. [*Cited on pages 29 and 35.*]

OLMSTEAD-RUMSEY, J. (2022): "Market Concentration and the Productivity Slowdown," *Working paper*. [*Cited on pages 4 and 34.*]

RAO, N. (2016): "Do tax credits stimulate R&D spending? The effect of the R&D tax credit in its first decade," *Journal of Public Economics*, 140, 1–12. [*Cited on page 68.*]

RESCHKE, B. P., P. AZOULAY, AND T. E. STUART (2018): "Status spillovers: The effect of status-conferring prizes on the allocation of attention," *Administrative Science Quarterly*, 63, 819–847. [*Cited on page 14.*]

ROACH, M. AND W. M. COHEN (2013): "Lens or prism? Patent citations as a measure of knowledge flows from public research," *Management Science*, 59, 504–525. [*Cited on page 9.*]

ROMER, P. M. (1990): "Endogenous technological change," *Journal of political Economy*, 98, S71–S102. [*Cited on page 4.*]

SAMPAT, B. AND H. L. WILLIAMS (2019): "How do patents affect follow-on innovation? Evidence from the human genome," *American Economic Review*, 109, 203–36. [*Cited on pages 2 and 24.*]

SPITALNIC, P., S. HEFFLER, B. DICKENSHEETS, AND M. KNIGHT (2016): "Hospital multifactor productivity: An updated presentation of two methodologies," *Office of the Actuary, Centers for Medicare & Medicaid Services, US Department of Health and Human Services*. [*Cited on page 19.*]

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, 19–50. [*Cited on pages 4 and 13.*]

WILLIAMS, H. (2012): "Innovation Inducement Prizes: Connecting Research to Policy," *Journal of Policy Analysis and Management*, 31, 752–776. [*Cited on page 14.*]

WILSON, D. J. (2009): "Beggar thy neighbor? The in-state, out-of-state, and aggregate effects of R&D tax credits," *The Review of Economics and Statistics*, 91, 431–436. [*Cited on page 68.*]

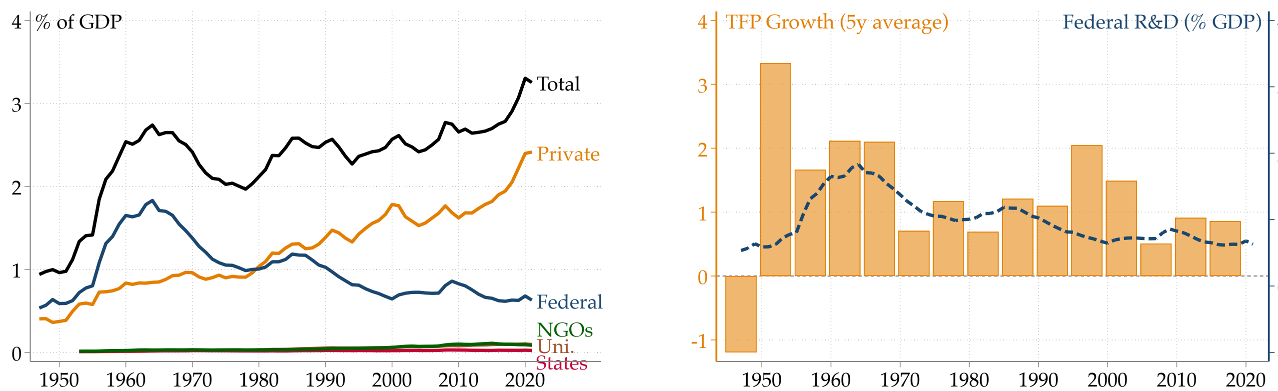
WILSON, D. J. *et al.* (2005): "The rise and spread of state R&D tax credits," *FRBSF Economic Letter*. [*Cited on page 67*.]

Appendix

TABLE OF CONTENTS

Appendix A. Historical trends in R&D funding

A.1. **Productivity growth and the funding of R&D in the United States.** The debate about the importance of public R&D spillovers is made more relevant by the fact that, in modern growth theory, spillovers play a critical role in driving productivity growth. Understanding how spillovers from private R&D differ from those of public R&D is therefore essential to assess the consequences of the secular decline in US public R&D as a share of GDP over the past 60 years (shown in Figure 8, left panel).[53] If public and private R&D differ in their ability to generate spillovers, then this large compositional shift in R&D should have important consequences for innovation and productivity growth.[54]



(A) R&D expenditures, by type of funder

(B) Aggregate TFP growth

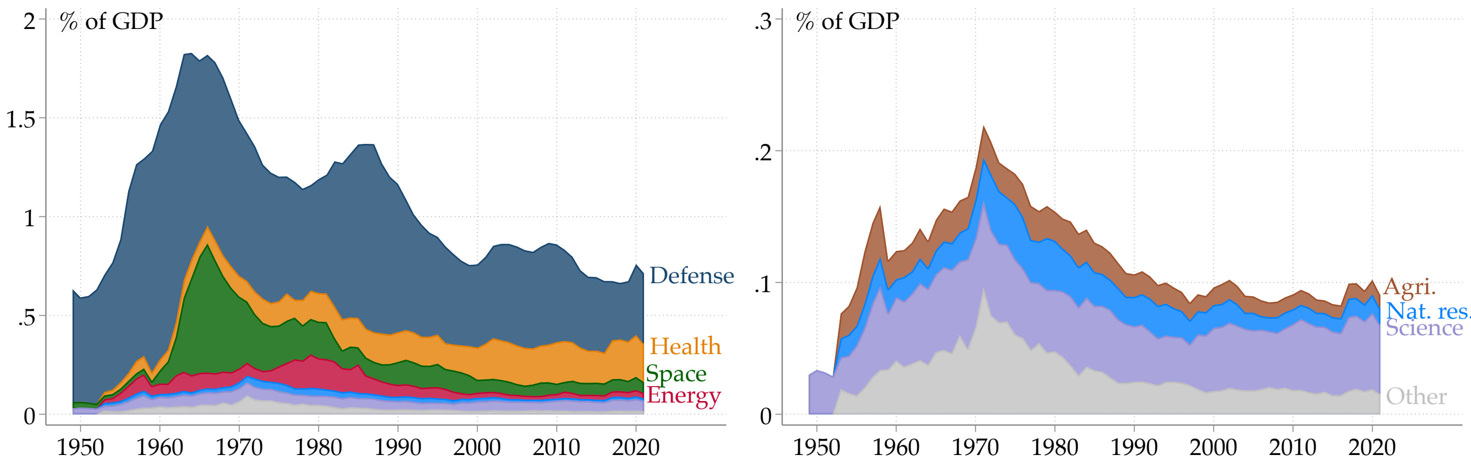
Figure 8. R&D funding and TFP growth in the US

Notes: Series on R&D expenditures come from the Bureau of Economic Analysis (pre-1953) and from the National Center for Science and Engineering Statistics, a National Science Foundation body (post-1953). Appendix A.2 breaks down federal R&D by departments and agencies. The aggregate TFP growth series comes from Bergeaud *et al.* (2016): each bar in the left panel is the geometric average of the aggregate TFP growth rate taken over five-year bins.

A.2. **Breakdown of public R&D funding over the past 70 years in the US.** Figure 9 shows the breakdown of federal R&D expenditures as a share of US GDP, across agencies. The left panel shows all agencies and the right panel focuses on the those with the smallest R&D expenditures.

[53]In 1960, federal R&D–which accounts for nearly all public R&D in the US–accounted for 1.7% of GDP. In contrast, it was just .7% in 2020. Over the same period, the GDP share of private R&D tripled from .8 to 2.4%. While federal R&D has declined as a share of US GDP, its amount has steadily risen: it went from $78 billion in 1960 to $148 billion in 2020, both expressed in 2020 dollars.

[54]Over the same period, aggregate Total Factor Productivity (TFP) growth decelerated from a high of 2.1% per year in the early 1960s to .9% in the late 2010s as can be seen in the right panel of Figure 8.
Many other countries have experienced similar declines in public R&D over the last 40 years. See Figure 11 in the Appendix.



(A) All funders

(B) Zooming in on the smallest agencies

FIGURE 9. Federal R&D expenditures, by department and agency

Notes: Time series come from the database of historical trends in federal R&D assembled by the American Association for the Advancement of Science. The agency funding the R&D is not necessarily performing the R&D.

A.3. **Public R&D funding: the US and the rest of the world.** The US government is not alone in investing in public R&D, and international spillovers from other countries may affect American firms' performance (Liu and Ma, 2023). However, the US appears to be the most important player when it comes to public R&D. The OECD provides data on government-funded R&D over the last 40 years: it shows that the US public R&D budget has been as large as the sum of all other OECD countries' public R&D budgets, from 1981 to 2022.[55]

Furthermore, the American economy relies relatively little on spillovers from other countries. In a recent working paper exploring cross-industry spillovers, Liu and Ma (2023) document that countries are heterogeneous in their degree of reliance on domestically produced knowledge. The US and Japan exhibit large shares of patent citations to domestically produced patents (around 70% for both countries) while countries like France and the United Kingdom have a majority of their patent citations directed toward international patents. Taking these citation patterns as indicators of knowledge spillovers, the authors conclude that the US is a large net exporter of knowledge to other countries.

Lastly, knowledge spillovers are usually very localized and do not travel far. A voluminous literature about knowledge spillovers started by Jaffe and Trajtenberg (1999) has documented that

[55]The other countries in the data are Australia, Austria, Belgium, Canada, Chile, Colombia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom. Budgets are expressed in 2015 dollars. The data is from the 'Gross domestic expenditure on R-D by sector of performance and source of funds' series.

they decay very rapidly with distance. When measured by patent citations, most spillovers occur in the immediate vicinity of where the knowledge was produced and do not travel much further than the region around a city. This effect is particularly true for more advanced, less codified knowledge.

These three facts lend support to the choice of this paper to focus on spillovers from US public R&D only. Including international spillovers could be an interesting extension of the present work. The most important reason why one would want to look into international spillovers is the recent rise of China's public R&D budget over the last 20 years. Indeed, the US budget was six times as big as the Chinese one in 2003, the first year when OECD data is available, but it is only 1.2 times as big in 2022.

A.4. **The (un)importance of R&D tax credits.** R&D tax credits are used in many countries to incentivize private R&D spending. This section assesses if the federal and local R&D credits available to US firms are likely to have fueled the rise in private R&D. Because of the limited generosity of the federal tax credit, its late introduction in 1981 and the unavailability of local state credits in some state, I conclude that it is unlikely that R&D tax credits are behind the secular rise in private R&D in the US.

Introduced in 1981 as part of the Economic Recovery Tax Act, the 'Credit for Increasing Research Activities' is the tax relief scheme used by the federal government to foster private R&D in the United States. It enables firms to claim a tax relief of up to 20% of R&D expenses (in excess of a base amount), provided the expenses satisfy eligibility criteria. Qualified research expenses include wages, material costs and rental cost of certain scientific property and equipment used in research. The two main components of the scheme are the Regular Research Credit (RRC), typically used by larger firms with a history of R&D, and the Alternative Simplified Credit (ASC), typically used by smaller and younger firms. In addition, firms can claim refunds on basic research expenses and energy research expenses. If a company's tax liability is insufficient to fully utilize the credit, the unused portion can be carried forward for up to 20 years. Additionally, since 2016, eligible start-ups have the option to apply a portion of their research credit, up to $250,000, against their payroll tax liability instead of their income tax liability. Wages paid to do in-house R&D constitute the largest expense eligible for the credit.

R&D tax credits are unlikely to have fueled a significant proportion of the secular increase in private R&D shown in figure 8a. Firstly, they have been introduced only in 1981, more than three decades after the rise in private R&D has been first recorded. Secondly, the American federal tax credit is not particularly generous compared to similar fiscal incentives in OECD countries (OECD,

2021) and it accounts for a small share of total private R&D.[56] To gauge the importance of federal tax credits in aggregate private R&D, figure 10 plots the total amount of tax credits claimed by businesses, as a share of GDP (data is only available from 1990 to 2013). In 2013, American corporations claimed only $11 billion in R&D tax credits. In contrast, total private R&D spending was $297 billion that year. R&D credits can thus hardly explain the large increase in private R&D.

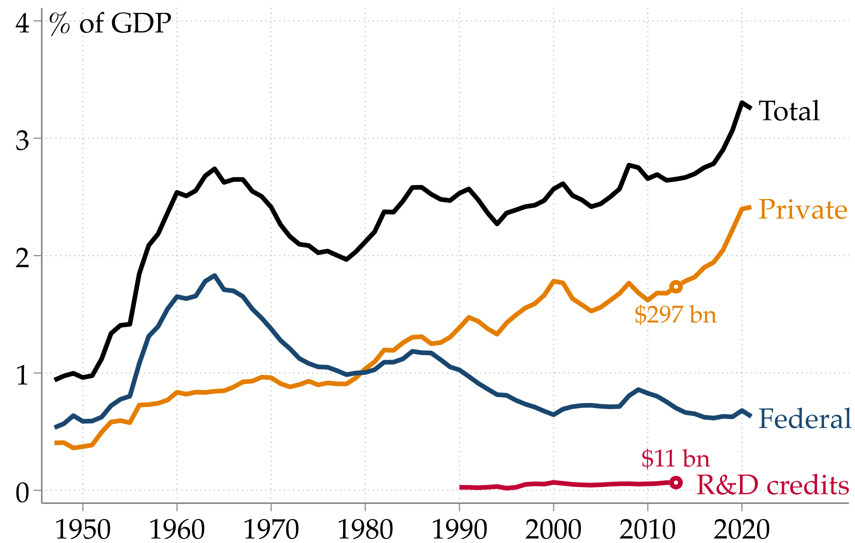


FIGURE 10. R&D tax credits and R&D expenses

Notes: Series on R&D expenditures come from the Bureau of Economic Analysis (pre-1953) and from the National Center for Science and Engineering Statistics, a National Science Foundation body (post-1953). Note that R&D expenditures by firms with fewer than 5 employees ('microbusinesses') are not counted in the NSF surveys on R&D spending before 2016. See NSF National Science Board (2022), footnote 5, p. 73. The inclusion of microbusiness R&D in total private R&D makes little difference: it accounted for only $4 to $5 billion in 2016 (out of $375 billion, *i.e.* 1.3%), year of its inclusion.
Data on tax credits claims come from the IRS's Statistics of Income – Corporation Research Credit webpage.

Federal tax credits are not the only fiscal incentives R&D-performing firms have access to; as many as 36 states had their own R&D credit scheme in 2023. It is however unlikely that state tax credit matter much for several reasons. The first is that state tax credit rate is typically lower than the federal credit rate (from 1% to 20% according to Wilson *et al.* (2005)). Secondly, not all states offer R&D tax credits and very few were offereing tax credits in the 1980, shortly after

[56]The OECD rates the US R&D tax credit as less generous than the average OECD R&D tax credit, with an implied subsidy rate of 7% compared to 20% for the average OECD country (OECD, 2021). The implied subsidy rate is calculated as $1 - B_{\text{index}}$ where B_{index} is the level of pre-tax profit a representative company needs to make to break even on a marginal, unitary outlay on R&D. In other words, a B_{index} of 100% means that firms need to generate one dollar of profit to break even after one dollar of R&D expense. In 2021, American firms needed to make $0.93 of profits to justify a marginal dollar of R&D. French and German firms, on the other hand, only needed to make $0.60 and $0.80 of profits, respectively, because the taxes and subsidies there are more advantageous for R&D performing firms.

the introduction of the federal tax credit. Until 1984, only Maryland had a state tax credit. The number of states with credit then gradually increased to reach 31 in 2005. Lastly, careful analysis of the aggregate effects of state R&D tax credits by Wilson (2009) find that increases in private R&D ascribed to state credits come almost entirely from drawing away R&D from other states, such that changes in tax credits essentially leave aggregate R&D spending unchanged. Most state schemes follow federal guidelines to determine what constitute a qualified research expense and how generous the state credit should be. While no database of state tax credits exists, one may look at California, the most R&D-intensive state in the United States, to evaluate how important state tax credits are for total private R&D investment. California introduced its owned tax credit in 1987, six years after the federal one was enacted. It covers R&D activities performed in California only and allows firm to reduce their tax liability by 15% to 24% of their R&D expenses. In 2014, Californian firms claimed $1.5 billion in research credit (Melass *et al.*, 2021). This represents 12% of the $12.6 billion claimed in *federal* R&D credits that year (Guenther (2022), table 3, p. 16). To put this number in perspective, private R&D in California accounts for one third of all private R&D in the US in 2019.[57] In other words, while Californian firms represent a third of all private R&D, they claimed an amount equivalent to roughly one tenth of federal credits in state credits. Given the unavailability of local R&D credits in some states, the delay in the introduction of local credits compared to federal credits and the Californian experience with local credits, making the assumption that local R&D credits are as important as federal tax credits is likely to yield an upper bound on the total amount of tax credits claimed by US firms. If one makes this assumption, total tax credits in 2013 amount to $22 billion (less than 5% of total R&D spending). Recent estimates of the elasticity of own-R&D spending to R&D tax credit suggest that $1 in credit leads to a $2 increase in R&D (Rao, 2016; Agrawal *et al.*, 2014; **?**). Using this elasticity and our upper bound estimate of $22 billion in tax credit, one can estimate the increase in private R&D due to state and federal credits as being $44 billion in 2013, or 13% of all private R&D. Arguably not a large share, even for an upper bound estimate. Furthermore, federal tax credits have remained flat through the period for which data is available, while private R&D has grown monotonically, further reducing the explanatory power of R&D credits as a driver of private R&D. For all these reasons, it seems unlikely that R&D credits are a major force behind the rise in private R&D.

Another worry one might have is that R&D tax credits are incentivizing firms to re-classify non-research expenses into research expenses. The existing set of papers quantifying the extent of reallocation is small, but their message is fairly consensual: there seems to be little reallocation of non-R&D expenses to R&D expenses following the introduction of tax credits. **?** use the introduction of a more advantageous tax regime in the UK aimed at increasing the innovation of small

[57]See this 2021 note by the State Science & Technology Institute (SSTI).

enterprise to evaluate the impact of R&D tax credits. They find that treated firms did not experience a decrease in the quality (citations) of the average patent after the introduction of the policy. This indirectly supports the idea that re-labeling of non-R&D expenses may not be severe. However, in an analysis of a Chinese R&D tax credits (China's InnoCom program), Chen *et al.* (2021) find that re-labeled expenses may account for a quarter of all of the change in R&D expenses. All in all, the evidence on R&D expenses re-labeling, while not exhaustive, suggests that re-labeling is a real, but not large margin of response of firms.

A.5. R&D budgets of US federal agencies. Panels A.7, A.8 and A.9 show the raw R&D budgets of the agencies I use in the construction of my SSIV instrument. Values are expressed in billions of 2020 dollars (deflated using the CPI from the Bureau of Labor Statistics).

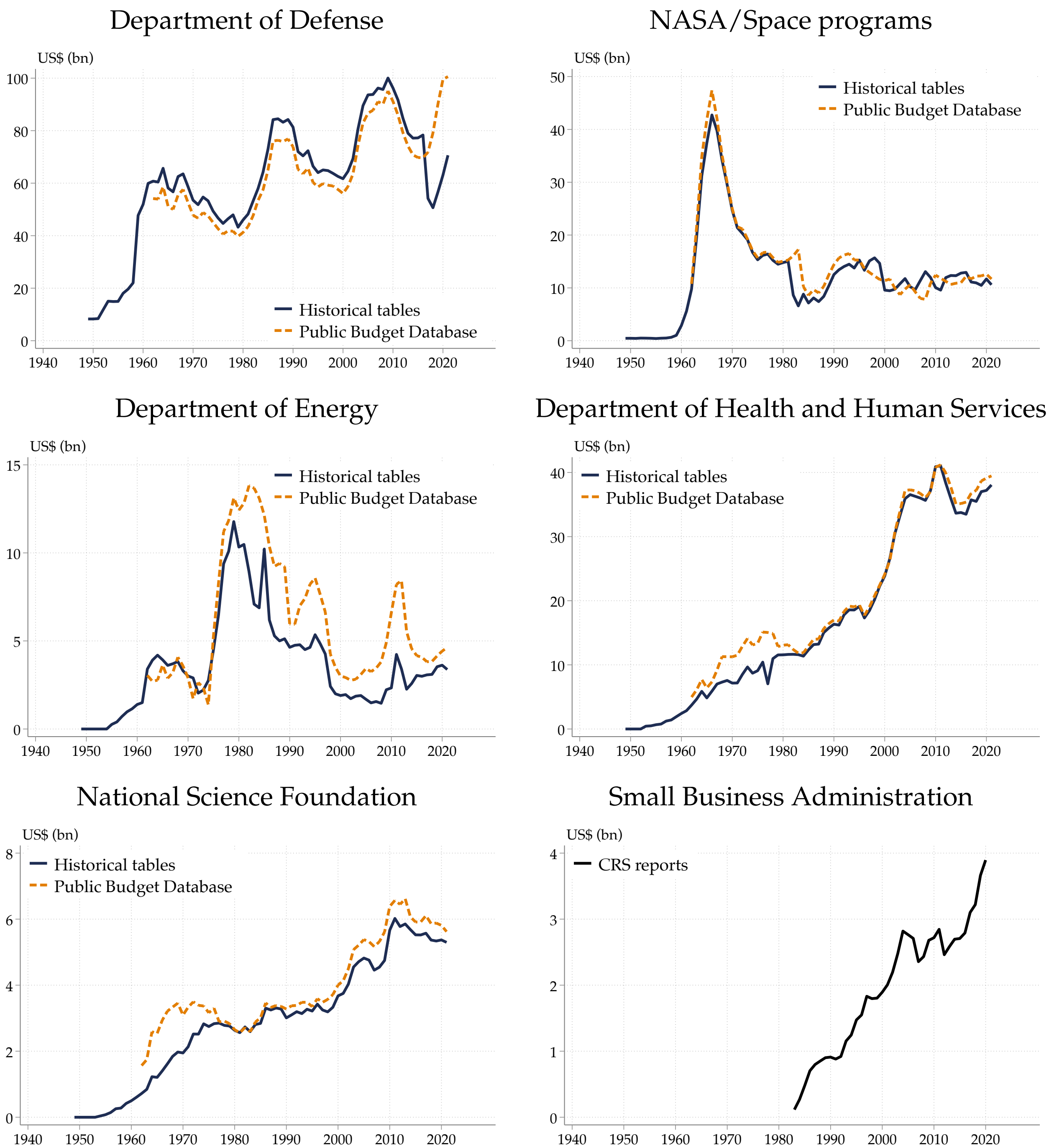


TABLE A.7. R&D budgets over time, federal agencies (1)

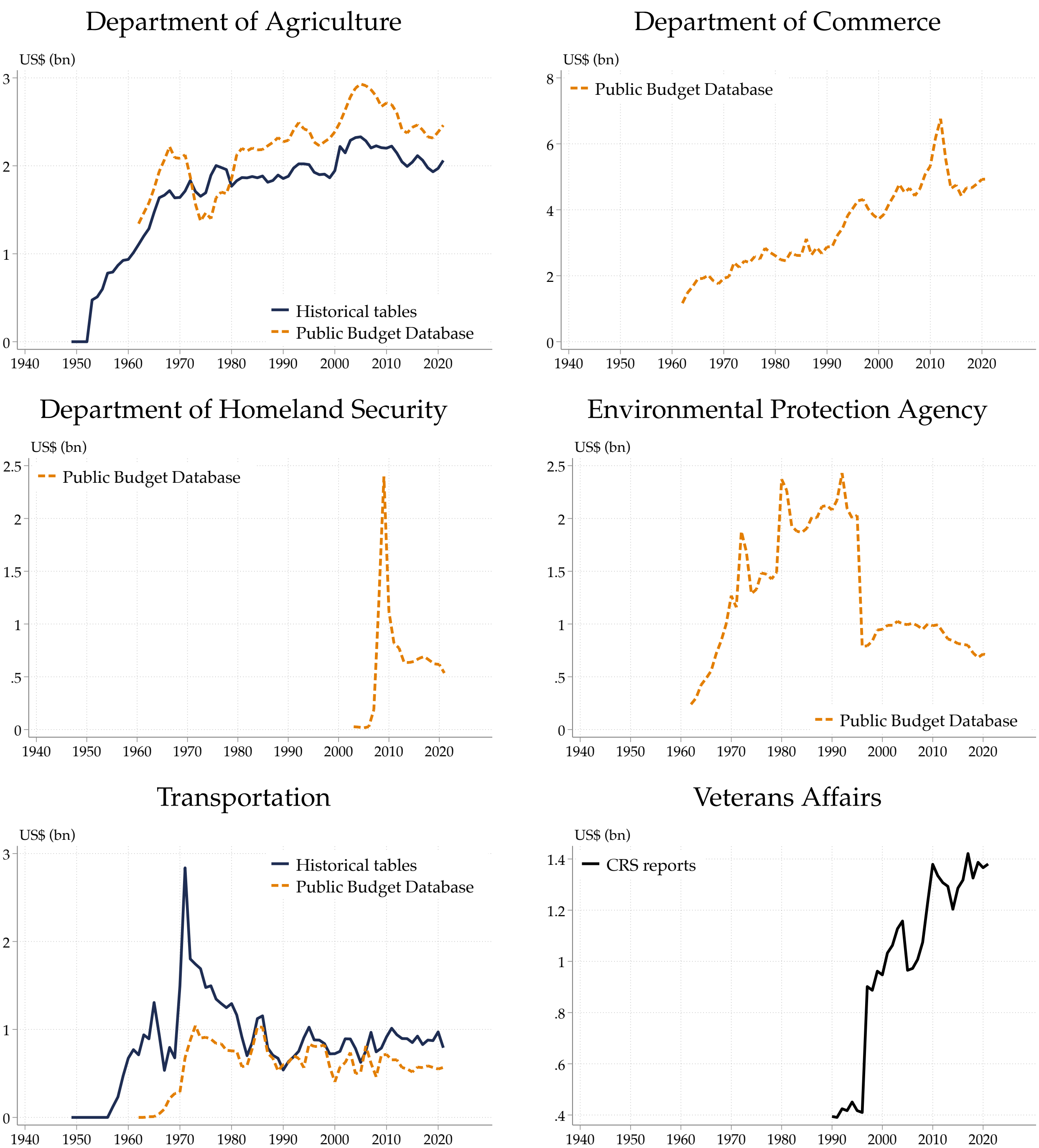


TABLE A.8. R&D budgets over time, federal agencies (2)

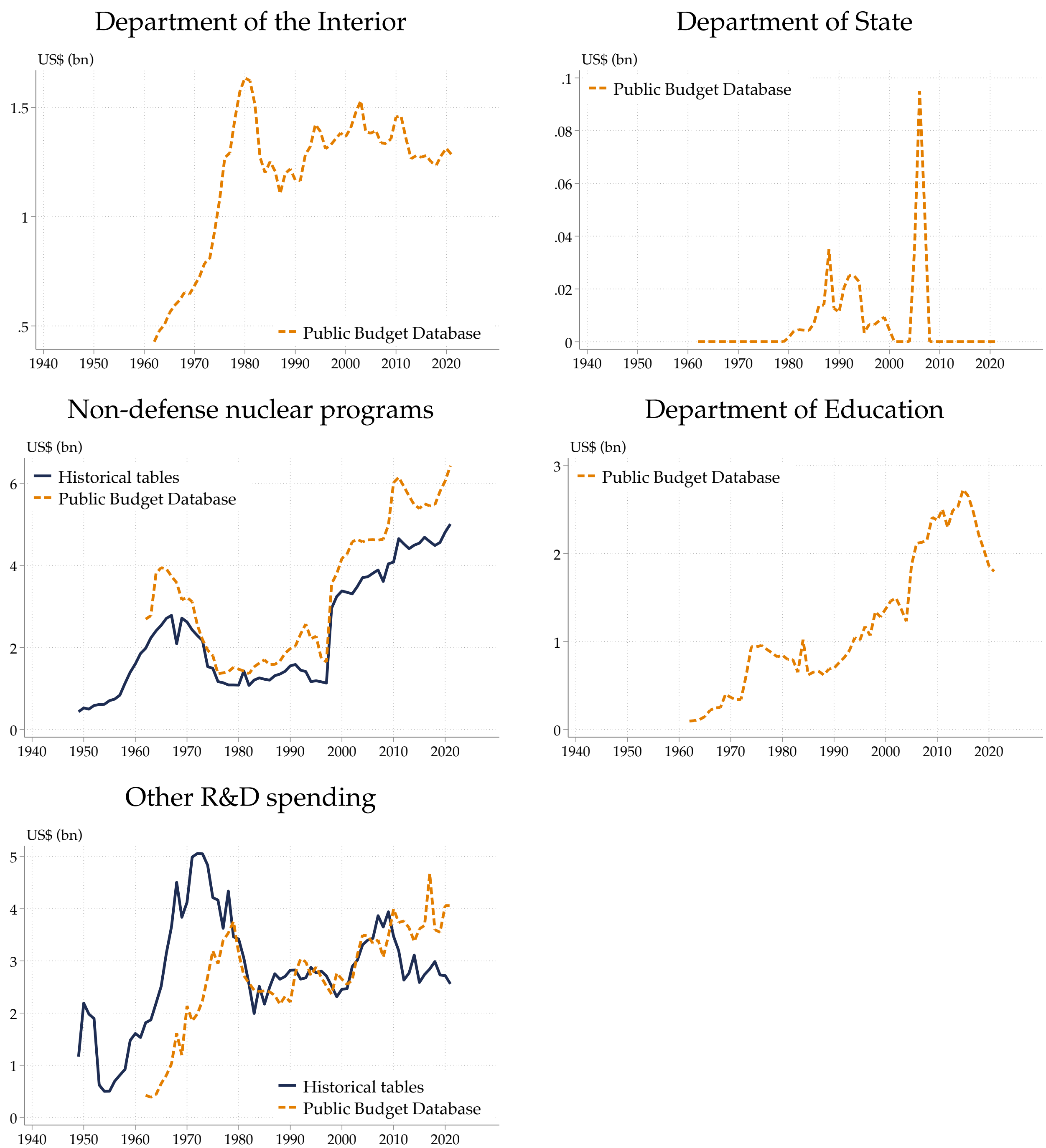


TABLE A.9. R&D budgets, federal agencies (3)

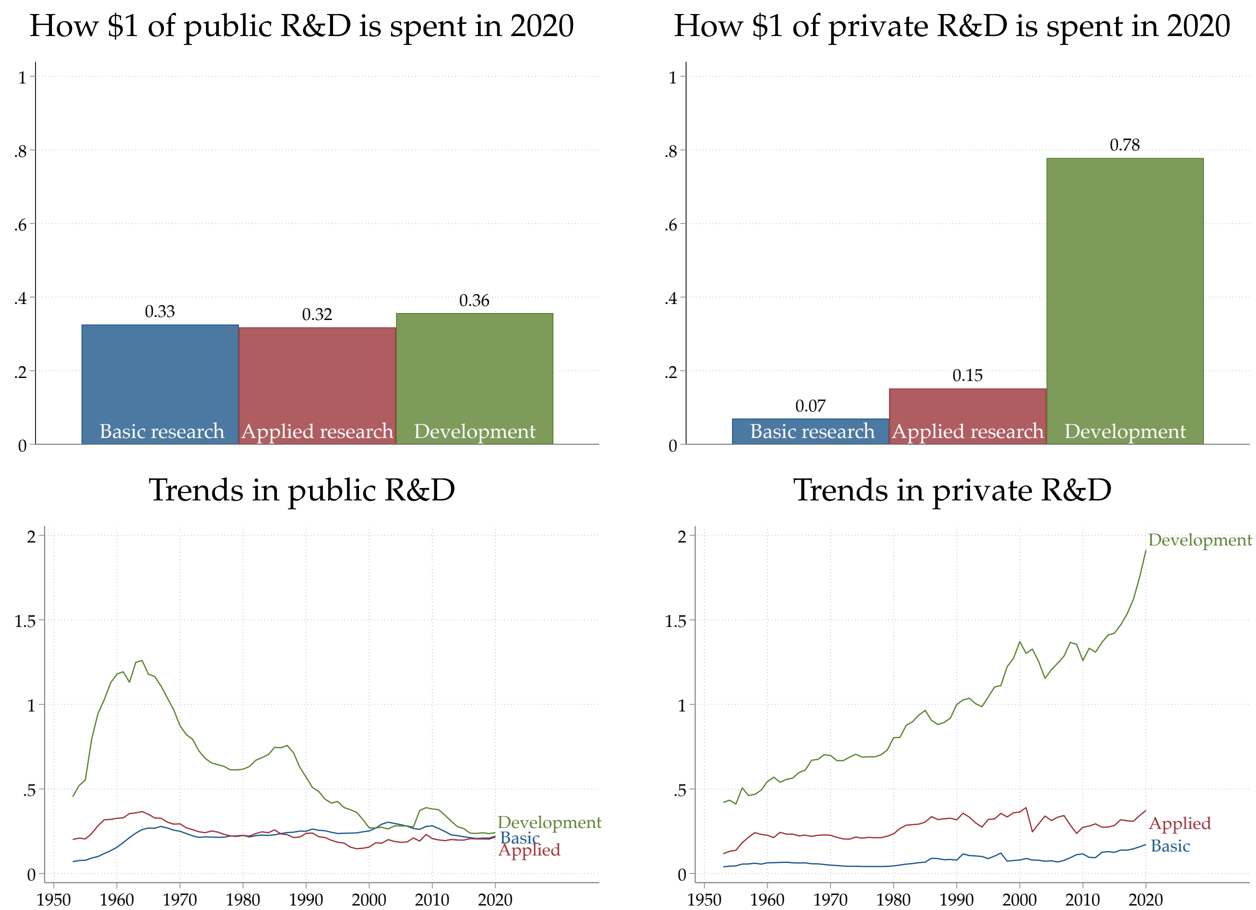
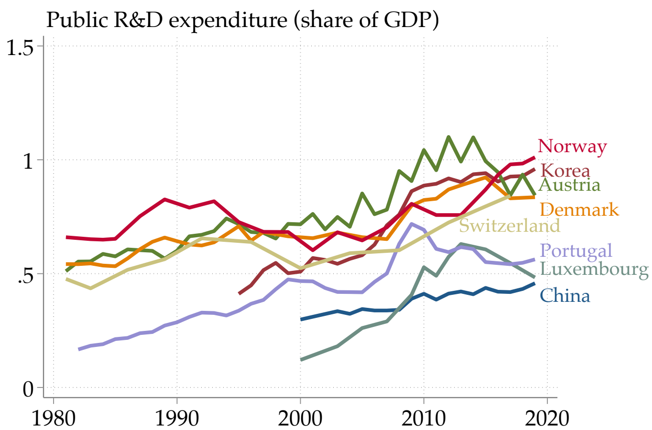


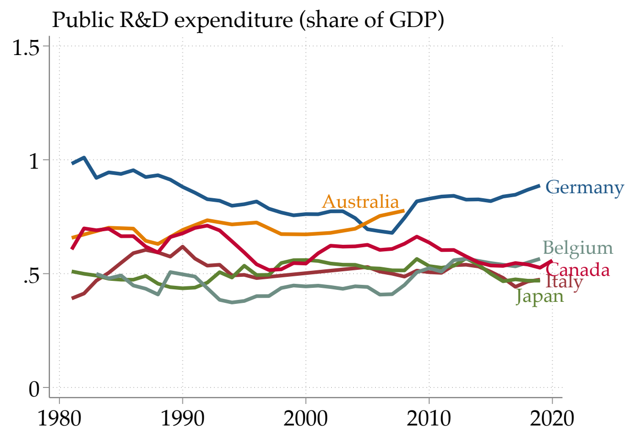
TABLE A.10. How public R&D differs from private R&D and trends over time

A.6. **Breakdown of public and private R&D.**

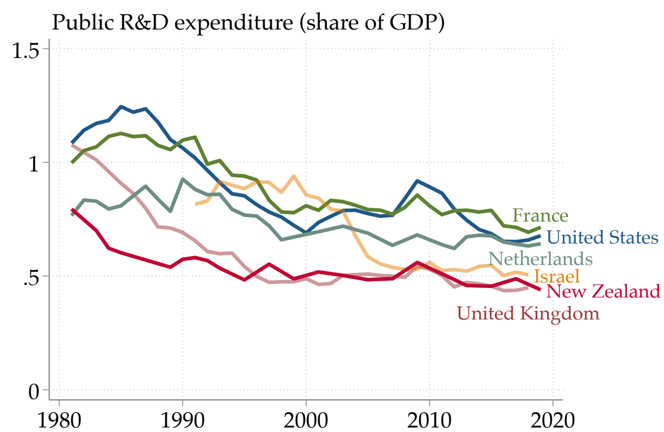
A.7. **Public R&D trends in other countries.** Outside of the US several advanced countries have also experienced a decline in public R&D as a share of GDP. The OECD provides data about government spending on R&D for several countries. The panels of Figure 11 show public R&D expenditures as a share of GDP for all countries for which data is available. Countries are classified in three groups depending on the growth trajectory of their public R&D as a share of GDP.



(A) Increasing



(B) Stable



(C) Decreasing

FIGURE 11. Historical public R&D trends in selected countries

Notes: Data come from the OECD, series 'Gross domestic expenditure on R&D by sector of performance and source of funds'. Available here.

Appendix B. Data appendix

B.1. **Other datasets of patents matched to firms.** [58]

There are two other main datasets of Compustat firms matched to patents: Arora *et al.* (2021b) and Kogan *et al.* (2017). Arora *et al.* (2021b) match USPTO patents to Compustat firms from 1985 to 2015, carefully reassigning patents from one firm to another after M&As, name changes and re-listings. Dyèvre and Seager (forthcoming) build on the work of Arora *et al.* (2021b), who themselves extend the matching efforts of Hall *et al.* (2001). We improve it in four ways. We first extend it temporally by matching USPTO patents to Compustat firms from 1950 to 2020, thereby covering the immediate postwar period which has experienced large swings in both federal budgets and patent production by agencies like NASA and the Department of Defense. We then improve the matching quality by manually reviewing matches between firm names in Compustat and assignee names in the USPTO datasets. Third, we add dynamic re-assignment events in the pre-1980 period. Finally, we add government interest tags to all patents.

We improve upon Kogan *et al.* (2017), which covers the period 1926-2022 in five ways: (i) by extending the coverage to 2020, (ii) by correcting many false positive matches in the original data due to the reliance of Kogan *et al.* (2017) on automated string cleaning algorithms, (iii) by adding government interest data, (iv) by using disambiguated patent data and most importantly, (v) by re-assigning patents after corporate events. While our dataset covers only three fourth of the period covered by KPSS's data, we are encompassing as many patents and a larger number of firms. Table B.11 summarizes the strengths of each dataset, including ours. The large coverage of firms over the 1950-2020 period and the dynamic nature of patent stocks make the DS dataset uniquely suited for the analyses performed in this paper.

B.2. Algorithm to match patents to Compustat firms. Due to the absence of firm identifiers that can join Compustat and the USPTO data, one has to rely on name matching to link firms to patent assignees. Our name matching algorithm, described in more details in Dyèvre and Seager (forthcoming), proceeds in four steps, and produces two datasets. The first dataset is called the *static* match. It assigns a firm in Compustat to each patent, at the time of filing. This dataset can be used to infer the flow of patents produced by a firm in a given year. The second dataset is a *dynamic* match. It provides associations between unique firm identifiers over ranges of years such that one can observe the evolution of a firm's patent stock over time.

To build this dataset, we combine data from nine sources: (i) patent data comes from PatentsView for patents filed between 1976 and 2020, (ii) patent data from 1950 to 1975 comes from Fleming *et al.* (2019), (iii) firm balance sheet data comes from Compustat North America, (iv) name changes

[58] In this section, 'we' refer to Arnaud Dyèvre and Oliver Seager, who have assembled the dataset used in this paper for another project.

	Coverage	Dynamic	Firms	Patents	Disambiguated
DS 2023 Used in this paper	1950-2020	✓	**9,961** unique GVKEYs	3.115m	PatentsView + Harmonization w/ FGLMY + Extensive manual checks
ABS 2021	1980-2015	✓	4,985 unique PERMNOs	1.349m	Extensive manual checks
KPSS 2023	**1926-2023**	No	8,547 unique PERMNOs	**3.160m**	Some manual checks
KPSS 2023 Restricted to 1950-2020	1950-2020	No	8,448 unique PERMNOs	2.918m	Some manual checks
NBER 2001	1963-1999	No	2,487 unique CUSIPs	0.835m	Automatic

TABLE B.11. Datasets of publicly-listed firms matched to patents

Notes: The numbers of patents and PERMNOs (unique firm identifier tied to a firm's stock) available in ABS 2021 are obtained from the `patent_1980_2015.dta` dataset from the authors (available here). The numbers for KPSS come from their `Match_patent_permco_permno_2022.csv` dataset (available here). The numbers for the NBER dataset come from the authors' `apat63_99.dta` dataset (available here).

and M&A data comes from the Center for Research and Security Prices (CRSP), (v) post-1985 corporate restructuring information comes from SDC Platinum, (vi) some data on firm ownership comes from Arora *et al.* (2021b), henceforth ABS, (vii) data from Wharton Research Data Services complements this information on corporate structure (because subsidiaries are listed in SEC 10-K filings), (viii) earlier data on corporate events comes from the list of acquisitions by publicly listed firms, from 1952 to 1963, compiled by Lev and Mandelker (1972) and finally (ix) a manually curated list of M&As, re-listings and spinoffs complements SDC Platinum (which starts in 1985) and Lev and Mandelker (1972) (which covers 1952-1963). With these datasets at hand, our merging effort proceeds in four steps. Our code is available in the project repository.

B.2.1. *Name cleaning.* Even within our two patent datasets, the same patent assignee may appear under different names because there are no unified reporting requirements. For instance, the technology firm IBM appears under 'I.B.M', 'IBM', 'International Business Machines', 'IBM Intellectual property' and many other names in the patent data. Furthermore, the FGLMY dataset contains a substantial amount of inaccurate firm names due to the authors' reliance on Optical Character Recognition (OCR) techniques to extract text from the patents PDFs. OCR is the only viable method to get patent information pre-1976, but further cleaning is required for this dataset. For instance, the machine-read text of a patent assignee field is 'Assignors to Reliance Electric and

Engineering of Ohio Application March 22 1947 Serial No. 736532' instead of 'Reliance Electric and Engineering'. We clean these firm names as best as we can before running the general name-cleaning algorithm on the combined patent datasets. To create a unique firm name for each relevant assignee, we homogenize names by removing leading and trailing white spaces, replacing non-standard characters such as 'é' or 'å' by standard ones, condensing acronyms such as 'Limited Liability Company' into 'LLC', replacing the names of large companies by a common name using a substring match (*e.g.* 'IBM' in 'IBM Intellectual Property') and finally removing all white spaces. As a result, 98.9% of all firm names in the patent datasets and 99.7% in the balance-sheet data are altered.

B.2.2. *Harmonization of firm names across patent datasets.* Even after cleaning firm names, we may still have discrepancies between the PatentsView and the FGLMY parts of the patent data. For instance, a firm may be reported as 'ABC Technologies' in FGLMY and 'ABC' in Compustat. In such cases, we leverage the joint coverage of both datasets from 1976 to 2017 and assign a new common name to assignees from PatentsView and FGLMY with significant overlap in patents. All assignees with significant overlap are subject to a careful manual review before being given a joint clean name. For the 250 firm names associated with the most patents, we also conduct online searches to find alternative names associated with the firm.

At the end of these three steps, we have 8,651,808 patents associated with 633,530 standardized firm names, from 1926 to 2020. We then proceed to match the assignee names to Compustat firm names

B.2.3. *Obtaining all the names under which a company trades.* A firm who files a patent under one name in a given year may not trade under the same name in another. Furthermore, patents filed by subsidiaries of a bigger firms need to be counted in the patent stock of the larger firm. The fourth step in our merging procedure consist in identifying all the names associated with each GVKEY-year pairs in Compustat. Following the methodology of Arora *et al.* (2021b), we fetch information on firm names from the CRSP Daily Stock file and CRSP-Compustat Linking Tables. 38% of all GVKEYs in our sample have at least two trading names over the 1950-2020 period. We then follow Bessen (2009) in attributing a patent to the highest level in a corporate structure by using subsidiary data from WRDS (which comes from SEC 10-K filings over the 1993-2019 period). We also rely on the work of ABS and Lev and Mandelker (1972) to get data on ownership and acquisitions of private subsidiaries, respectively. Finally, we add corporate events coming from a manually curated list covering the period from 1950 to 1980. All steps are subject to careful manual checks on the names of firms and the validity of the corporate events we identified.

B.2.4. *Dynamic match.* To then assign a patent to all the GVKEYs it is linked to, we fetch data on mergers, acquisitions, re-listings and spinoffs (henceforth 'corporate events') from four sources.

First, SDC Platinum provides 414 corporate events, from 1985 to 2020. Then, the CRSP-to-Compustat crosswalk provides an additional 570 corporate events over the whole period covered by Compustat. Third, we manually search for corporate events when we observe several GVKEYs associated with one standardized name. This step yields an additional 296 corporate events. Lastly, we review several lists of high-value M&A activity to complete the list of corporate events from 1950 to 1989 (a period with little to no coverage by SDC Platinum). This last step adds 700 additional corporate events.

B.3. **Detailed data description.**

Firms. I select companies headquartered in the US or Canada over 1950-2020. Nominal values are deflated using the CPI from the Bureau of Labor Statistics.

Patents. Patensview considerably improves upon previous disambiguation efforts by using hierarchical agglomerative clustering–a machine learning algorithm–to group differently spelled assignees into relevant categories (Monath *et al.*, 2021).[59]

Government interest. Both cases are identified separately. For direct assignees, I use the classification of Patentsview and Fleming *et al.* (2019) of assignees as government entities.[60] When necessary, I aggregate assignees to the highest level using the hierarchical table of government entities provided by PatensView[61] so that patents assigned to agencies like DARPA are aggregated up to the level of the Department of Defense for instance. This step ensures that the source of variation of federal budget funding is at the same level as the variation in patent production.

Patent examiner scores. The American Inventors Protection Act (AIPA) of 1999 mandates the public disclosure of most USPTO patent applications filed on or after November 29, 2000, regardless of whether the patents are eventually granted. Such applications are published in the public record

[59] Previous disambiguation efforts typically rely on 'edit-distance' techniques that assign a percentage of similarity between two strings based on how many characters need to be changed to transform one string into the other. For instance, an edit-distance procedure would assign high similarity scores to long assignee names with many characters in common such as 'The United States of America as represented by the secretary of the Navy' and 'The United States of America as represented by the secretary of the Army'. Such conflations would be problematic when assigning patents to government agencies. Conversely, assignees values 'I.B.M.' and 'International Business Machines' would not be paired. This type of false negative is the main reason behind my improvement over (Kogan *et al.*, 2017).

[60] It is common for government agencies to be assigned patents, even those producing innovations with a strategic interest.

[61] Table `g_gov_interest.tsv` provided by PatentsView

within 18 months of the filing date, with few exception such as applications which are national security classified or which are explicitly asked not to be published by the applicant.[62] The 2021 version of PatEx includes information on over 12.5 million non-provisional and provisional USPTO patent applications that are publicly viewable, as well as more than 1 million Patent Cooperation Treaty (PCT) applications. The data used for this version of PatEx was obtained by OCE from the Patent Examination Data System (PEDS) in June 2022. Coverage of patent applications is most reliable from December 2000 onward, when the AIPA enters in force: 83% of all post-AIPA applications are available in PatEx. Pre-AIPA coverage is only slightly less comprehensive, with three quarters of applications available (Graham *et al.*, 2018).

R&D Budgets. For agencies with no R&D budgets reported in these tables like the Department for Veterans Affairs, I recover their historical budgets from The first is the White House's website where R&D spending by agencies over the 1962-2022 period is reported in statistical tables.[63] The second is the official 2013 federal budget documents by the Office of Management and Budget which contained detailed accounts of expenditures by agencies from 1940 onward. I manually enter these numbers and, when missing, estimate R&D spending by scaling agencies' total budgets by the share of R&D in the federal government's total budget.

Other patent-related datasets. Dates of creation of technological fields come from data available on the USPTO website about the years of introduction of new USPC classes,[64] and patents disruptiveness scores come from Kelly *et al.* (2021).[65]

B.4. **Using patents to measure innovation and spillovers.** Patent documents contain detailed information about an innovation, its inventors, its assignees, and its technological content. The main limitation to the use of patents to measure innovation is that not all innovations are patented, either because the innovation does not meet one of the three main criteria for being protected by a patent (usefulness, novelty and non-obviousness) or because the invention is better protected by alternative means such as secrecy. However, there is a broad consensus that patent counts are a good, if noisy, indicator of the innovativeness of an inventor, a firm, a city or a country.

[62] Applications that are not published 18 months after filing may be published 60 months after filing instead. Although some US patent applications may choose to opt out of publication, according to Graham and Hegde's 2013 study, only around 8 percent of US applications have chosen to do so for pre-grant secrecy of patent applications.

[63] www.whitehouse.gov/omb/budget/historical-tables/, table 9.8.

[64] Raw data stored at the following link arnauddyevre.com/files/USPC_classes_years_established.pdf. Csv file available at arnauddyevre.com/files/timeline_detail_classes.csv

[65] Data made available by the authors at dimitris-papanikolaou.github.io/website/

Patent counts are typically strongly correlated with measures of inputs into the innovation process such as R&D expenses or the number of researchers in a firm. There is also evidence that a firm's patent count is positively associated with many metrics of firm performance. For instance, the patent yield of R&D expenses (measured as the ratio of patents to R&D expenditure) is positively associated with a firm's Tobin q (Hall *et al.*, 2005).

Moreover, citations are a good indicator of the economic value of a patent, a evidenced by the positive association between the average citation count received by patents and the filing firm's Tobin's q Hall *et al.* (2005). They are also good proxies for the technological value of patents as: expert valuations of the merits of patents correlate positively with their citation counts (Albert *et al.*, 1991) and patents who are 'Hall of Fame' or identified by patent offices as being important are highly cited (Narin, 1995). In contrast, patents expertly idenified as futile receive fewer citations (Czarnitzki *et al.*, 2011). Benson and Magee (2015) also show that the citation counts of patents in some technological domains is positively associated with the rate of progress (the reduction in costs for instance) in these domains. When studying the strategic decisions of firms of different sizes to expose themselves to outward spillover, Crescenzi *et al.* (2022) find that the quantity of citations to foreign firms in a region is a signal of spillovers that is correlated with other signals such as inventor movements between firms and joint patenting.

See Jaffe and De Rassenfosse (2017) for a recent overview of best practices.

B.5. **Shares of proximity in technology space over time.** The shares of proximity s_{ift} and s_{iat} used in my empirical exercises are time-varying but they appear to be extremely sticky in the data. Figure 12 shows the correlation between shares of exposure to federal agencies in one five-year period (on the x-axis) and shares of exposure in the next five-year period (y-axis). All shares are very close to the 45 degree line. Shares in future periods are larger due to the increase in the number of federal agencies over time.

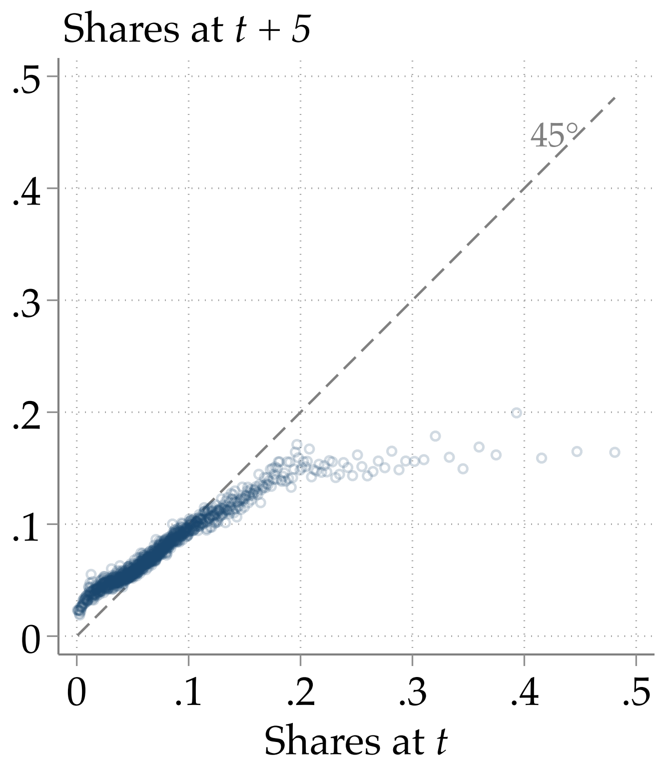


FIGURE 12. Stability of shares of exposures to public R&D

Notes: The figure plots a binscatter of firm-to-agency exposure shares, from each 5-year period to the next. Each dot represent approximately 170 firm × period observations. The plot uses 1,000 bins, defined at t, to facilitates legibility. The correlation between shares over time is 0.61. The top 3 agencies with the highest average firm exposures are the Department of Defense (firms exposed to the DoD have a 17.8% exposure on average), NASA (13.7%), the Department of Agriculture and the Department of Energy (both at 10.8%).

Appendix C. Additional results on public & private R&D patents

C.1. Historical USPC classes. Figure 13 shows the cumulative shares of USPC patent classes is use over time. The blue time series uses the date of introduction of classes while the red one uses the data of the first patent in the new classes. Because patents are *ex post* re-classified into the most relevant patent class, the blue time series first order stochastically dominates the red one. See Lafond and Kim (2019) for a detailed history of the USPTO classification system.

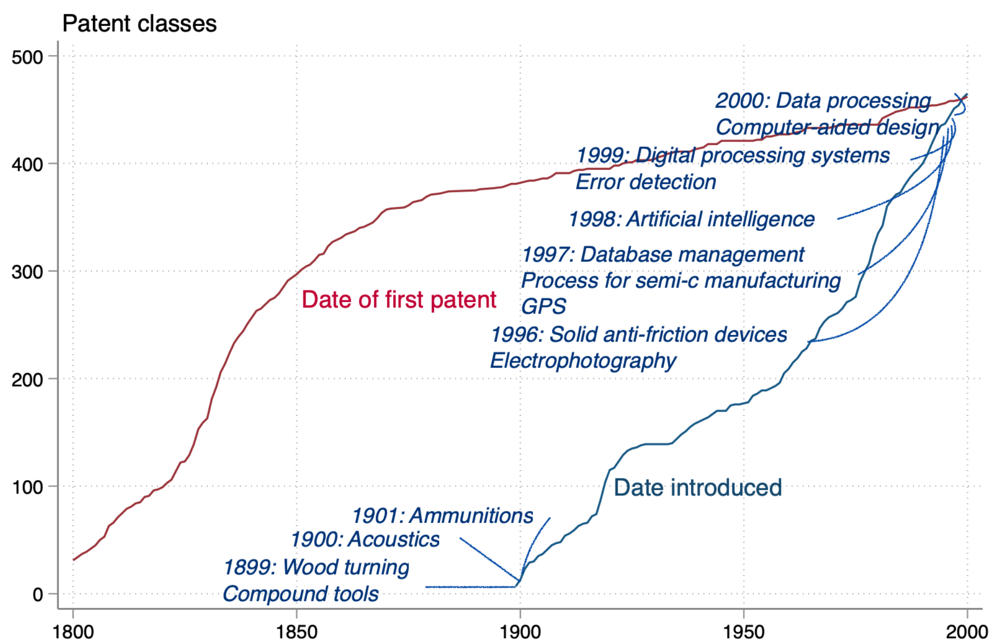


FIGURE 13. Timeline of the introduction of new USPC patent classes

C.2. All results - publicly-funded vs. privately-funded patents.

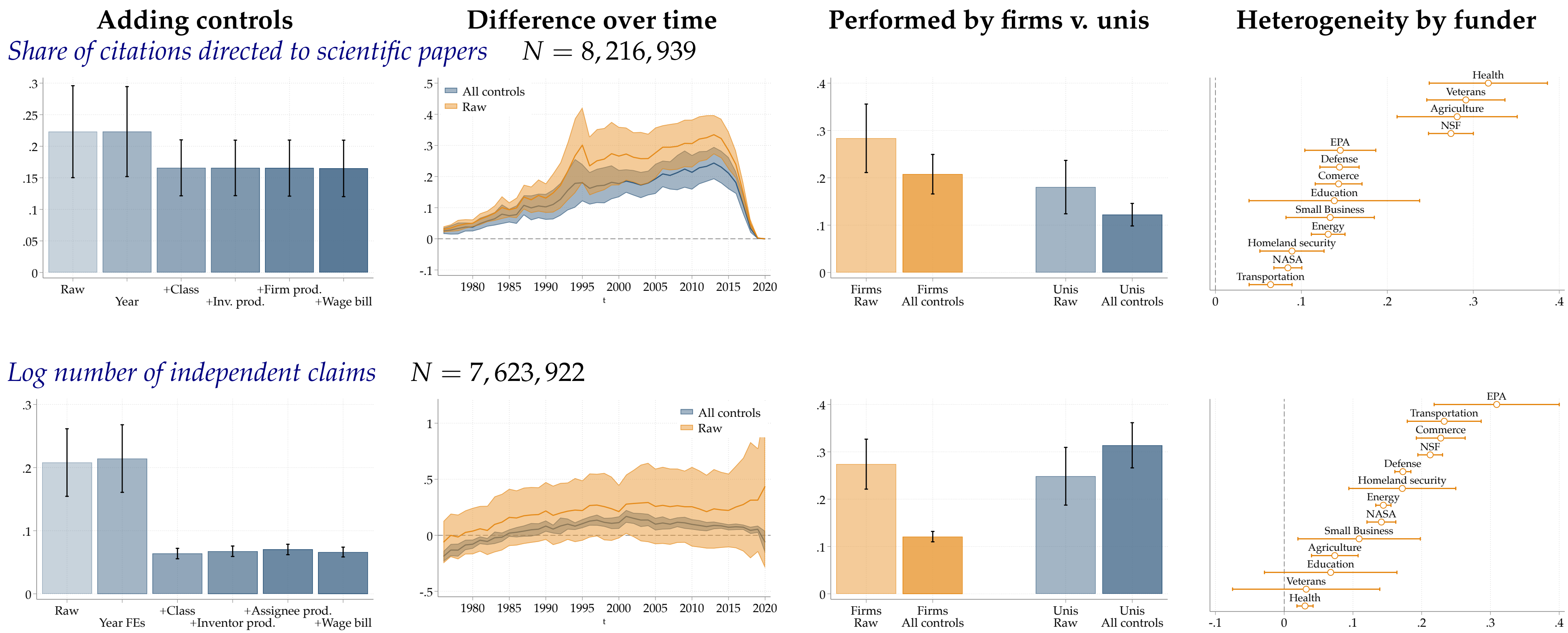


TABLE C.12. Fact 1 – Publicly-funded patents are more fundamentals (all results)

Notes: The unit of analysis is a patent. Coefficients and 95% confidence intervals come from a regression of an outcome of interest (y_i) on a dummy equal to one if the innovation protected by the patent benefited from public funding. Formally: $y_i = \alpha + \beta \times \mathbb{1}[\text{patent } i \text{ is publicly-funded}] + \mathbf{X}_i\gamma + \varepsilon_i$. Standard errors are clustered at the class and year levels. Graphs in the first column show how β varies when successively more exhaustive arrays of controls are used. Graphs in the second column report β coefficients for different years. Graphs in the third column show how the β coefficient varies within *performers* of R&D: universities or firms. The last graphs report coefficient heterogeneity across R&D funders.

TABLE C.13. Fact 2 – Publicly-funded patents are more impactful (all results)

Notes: The unit of analysis is a patent. Coefficients and 95% confidence intervals come from a regression of an outcome of interest (y_i) on a dummy equal to one if the innovation protected by the patent benefited from public funding. Formally: $y_i = \alpha + \beta \times \mathbb{1}[\text{patent } i \text{ is publicly-funded}] + \mathbf{X}_i\gamma + \varepsilon_i$. Standard errors are clustered at the class and year levels. Graphs in the first column show how β varies when successively more exhaustive arrays of controls are used. Graphs in the second column report β coefficients for different years. Graphs in the third column show how the β coefficient varies within *performers* of R&D: universities or firms. The last graphs report coefficient heterogeneity across R&D funders.

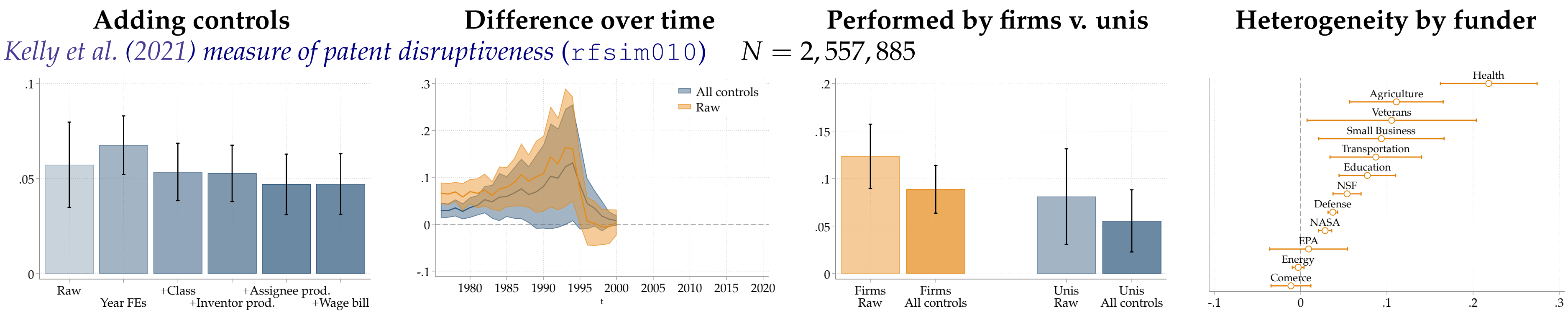


TABLE C.14. Fact 2 (continued) – Publicly-funded patents are more impactful (all results)

Notes: The unit of analysis is a patent. Coefficients and 95% confidence intervals come from a regression of an outcome of interest (y_i) on a dummy equal to one if the innovation protected by the patent benefited from public funding. Formally: $y_i = \alpha + \beta \times \mathbb{1}[\text{patent } i \text{ is publicly-funded}] + \mathbf{X}_i\gamma + \varepsilon_i$. Standard errors are clustered at the class and year levels. Graphs in the first column show how β varies when successively more exhaustive arrays of controls are used. Graphs in the second column report β coefficients for different years. Graphs in the third column show how the β coefficient varies within *performers* of R&D: universities or firms. The last graphs report coefficient heterogeneity across R&D funders.

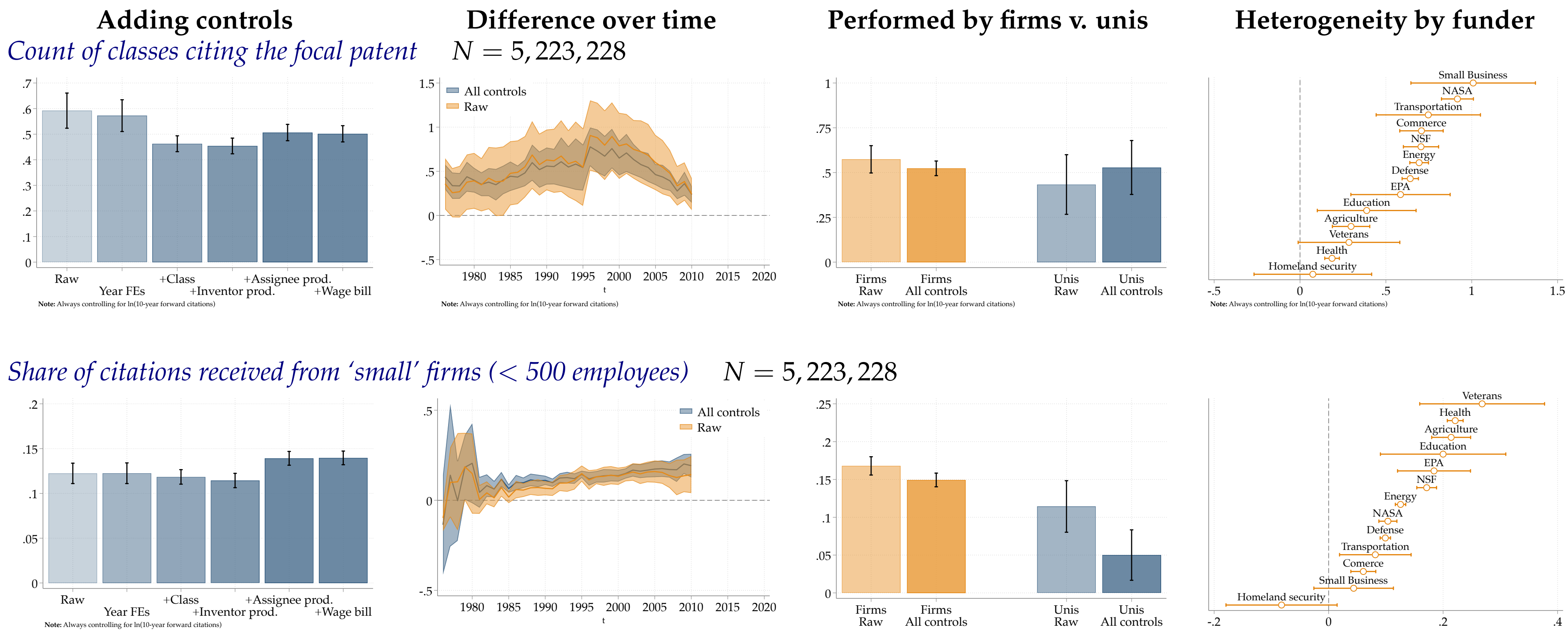


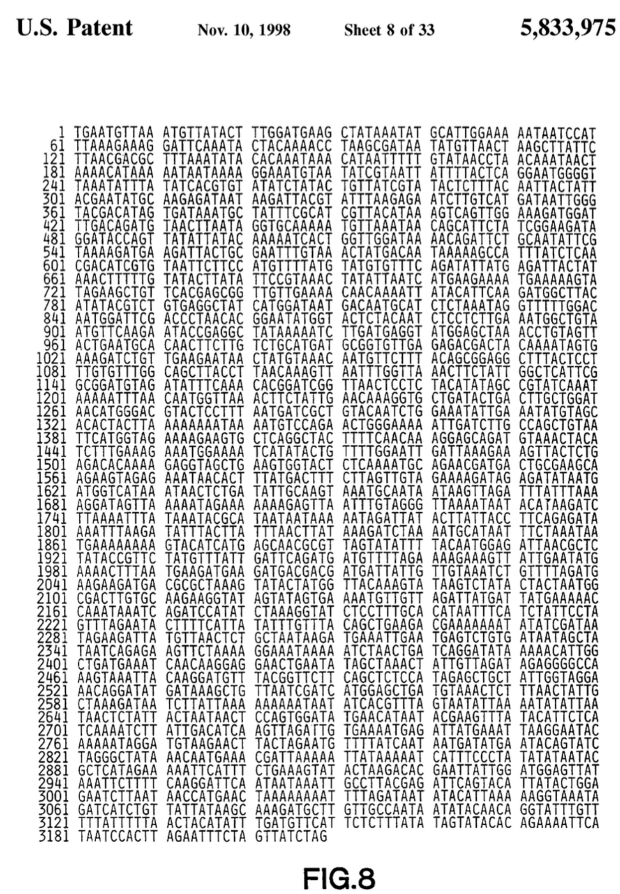
TABLE C.15. Fact 3 – Publicly-funded patents generate more spillovers, especially to small firms (all results)

Notes: The unit of analysis is a patent. Coefficients and 95% confidence intervals come from a regression of an outcome of interest (y_i) on a dummy equal to one if the innovation protected by the patent benefited from public funding. Formally: $y_i = \alpha + \beta \times \mathbb{1}[\text{patent } i \text{ is publicly-funded}] + \mathbf{X}_i\gamma + \varepsilon_i$. Standard errors are clustered at the class and year levels. Graphs in the first column show how β varies when successively more exhaustive arrays of controls are used. Graphs in the second column report β coefficients for different years. Graphs in the third column show how the β coefficient varies within *performers* of R&D: universities or firms. The last graphs report coefficient heterogeneity across R&D funders.

C.3. **Some case studies.** To fix ideas, and to better understand which publicly-funded patents do well across the outcome variables used in section 3, it is informative to study a few patents in more details. I present here three case studies of government-supported technologies. The first case study describes the government-supported patent that relies most heavily on science in my sample. The second is the one that is most 'ahead of its time', and the last one is the government-supported patent cited by the largest number of patent classes.

C.3.1. *Case study 1 – An innovation in immunotherapy that relies on medical science.* In my sample of patents, patent number 5,833,975 is the one with the highest share of citations to scientific articles. Only five of its citations are directed to previous patents and the remaining 492 are directed to scientific papers (99% of the total).

The process protected by this patent is one whereby medical researchers can modify poxviruses in order to use them as insertion and expression vehicles for genes in a host body. These genes are used in immunization processes; they enable the expression of an 'antigenic protein' that can induce an immunological response in the host. An important application of this technology is the development of immunotherapy for patients treated for cancers. Figure 14, taken from the patent, shows one of several DNA sequences of genes that can be expressed by the modified poxviruses.



U.S. Patent Nov. 10, 1998 Sheet 8 of 33 5,833,975

```
   1 TGAATGTTAA ATGTTATACT TTGGATGAAG CTATAAATAT GCATTGGAAA AATAATCCAT
  61 TTAAAGAAAG GATTCAAATA CTACAAAACC TAAGCGATAA TATGTTAACT AAGCTTATTC
 121 TTAACGACGC TTTAAATATA CACAAATAAA CATAATTTTT GTATAACCTA ACAAATAACT
 181 AAAACATAAA AATAATAAAA GGAAATGTAA TATCGTAATT ATTTTACTCA GGAATGGGGT
 241 TAAATATTTA TATCACGTGT ATATCTATAC TGTTATCGTA TACTCTTTAC AATTACTATT
 301 ACGAATATGC AAGAGATAAT AAGATTACGT ATTTAAGAGA ATCTTGTCAT GATAATTGGG
 361 TACGACATAG TGATAAATGC TATTTCGCAT CGTTACATAA AGTCAGTTGG AAAGATGGAT
 421 TTGACAGATG TAACTTAATA GGTGCAAAAA TGTTAAATAA CAGCATTCTA TCGGAAGATA
 481 GGATACCAGT TATATTATAC AAAAATCACT GGTTGGATAA AACAGATTCT GCAATATTCG
 541 TAAAAGATGA AGATTACTGC GAATTTGTAA ACTATGACAA TAAAAAGCCA TTTATCTCAA
 601 CGACATCGTG TAATTCTTCC ATGTTTTATG TATGTGTTTC AGATATTATG AGATTACTAT
 661 AAACTTTTTG TATACTTATA TTCCGTAAAC TATATTAATC ATGAAGAAAA TGAAAAAGTA
 721 TAGAAGCTGT TCACGAGCGG TTGTTGAAAA CAACAAAATT ATACATTCAA GATGGCTTAC
 781 ATATACGTCT GTGAGGCTAT CATGGATAAT GACAATGCAT CTCTAAATAG GTTTTTGGAC
 841 AATGGATTCG ACCCTAACAC GGAATATGGT ACTCTACAAT CTCCTCTTGA AATGGCTGTA
 901 ATGTTCAAGA ATACCGAGGC TATAAAAATC TTGATGAGGT ATGGAGCTAA ACCTGTAGTT
 961 ACTGAATGCA CAACTTCTTG TCTGCATGAT GCGGTGTTGA GAGACGACTA CAAAATAGTG
1021 AAAGATCTGT TGAAGAATAA CTATGTAAAC AATGTTCTTT ACAGCGGAGG CTTTACTCCT
1081 TTGTGTTTGG CAGCTTACCT TAACAAAGTT AATTTGGTTA AACTTCTATT GGCTCATTCG
1141 GCGGATGTAG ATATTTCAAA CACGGATCGG TTAACTCCTC TACATATAGC CGTATCAAAT
1201 AAAAATTTAA CAATGGTTAA ACTTCTATTG AACAAAGGTG CTGATACTGA CTTGCTGGAT
1261 AACATGGGAC GTACTCCTTT AATGATCGCT GTACAATCTG GAAATATTGA AATATGTAGC
1321 ACACTACTTA AAAAAAATAA AATGTCCAGA ACTGGGAAAA ATTGATCTTG CCAGCTGTAA
1381 TTCATGGTAG AAAAGAAGTG CTCAGGCTAC TTTTCAACAA AGGAGCAGAT GTAAACTACA
1441 TCTTTGAAAG AAATGGAAAA TCATATACTG TTTTGGAATT GATTAAAGAA AGTTACTCTG
1501 AGACACAAAA GAGGTAGCTG AAGTGGTACT CTCAAAATGC AGAACGATGA CTGCGAAGCA
1561 AGAAGTAGAG AAATAACACT TTATGACTTT CTTAGTTGTA GAAAAGATAG AGATATAATG
1621 ATGGTCATAA ATAACTCTGA TATTGCAAGT AAATGCAATA ATAAGTTAGA TTTATTTAAA
1681 AGGATAGTTA AAAATAGAAA AAAAGAGTTA ATTTGTAGGG TTAAAATAAT ACATAAGATC
1741 TTAAAATTTA TAAATACGCA TAATAATAAA AATAGATTAT ACTTATTACC TTCAGAGATA
1801 AAATTTAAGA TATTTACTTA TTTAACTTAT AAAGATCTAA AATGCATAAT TTCTAAATAA
1861 TGAAAAAAAA GTACATCATG AGCAACGCGT TAGTATATTT TACAATGGAG ATTAACGCTC
1921 TATACCGTTC TATGTTTATT GATTCAGATG ATGTTTTAGA AAAGAAAGTT ATTGAATATG
1981 AAAACTTTAA TGAAGATGAA GATGACGACG ATGATTATTG TTGTAAATCT GTTTTAGATG
2041 AAGAAGATGA CGCGCTAAAG TATACTATGG TTACAAAGTA TAAGTCTATA CTACTAATGG
2101 CGACTTGTGC AAGAAGGTAT AGTATAGTGA AAATGTTGTT AGATTATGAT TATGAAAAAC
2161 CAAATAAATC AGATCCATAT CTAAAGGTAT CTCCTTTGCA CATAATTTCA TCTATTCCTA
2221 GTTTAGAATA CTTTTCATTA TATTTGTTTA CAGCTGAAGA CGAAAAAAAT ATATCGATAA
2281 TAGAAGATTA TGTTAACTCT GCTAATAAGA TGAAATTGAA TGAGTCTGTG ATAATAGCTA
2341 TAATCAGAGA AGTTCTAAAA GGAAATAAAA ATCTAACTGA TCAGGATATA AAAACATTGG
2401 CTGATGAAAT CAACAAGGAG GAACTGAATA TAGCTAAACT ATTGTTAGAT AGAGGGGCCA
2461 AAGTAAATTA CAAGGATGTT TACGGTTCTT CAGCTCTCCA TAGAGCTGCT ATTGGTAGGA
2521 AACAGGATAT GATAAAGCTG TTAATCGATC ATGGAGCTGA TGTAAACTCT TTAACTATTG
2581 CTAAAGATAA TCTTATTAAA AAAAAATAAT ATCACGTTTA GTAATATTAA AATATATTAA
2641 TAACTCTATT ACTAATAACT CCAGTGGATA TGAACATAAT ACGAAGTTTA TACATTCTCA
2701 TCAAAATCTT ATTGACATCA AGTTAGATTG TGAAAATGAG ATTATGAAAT TAAGGAATAC
2761 AAAAATAGGA TGTAAGAACT TACTAGAATG TTTTATCAAT AATGATATGA ATACAGTATC
2821 TAGGGCTATA AACAATGAAA CGATTAAAAA TTATAAAAAT CATTTCCCTA TATATAATAC
2881 GCTCATAGAA AAATTCATTT CTGAAAGTAT ACTAAGACAC GAATTATTGG ATGGAGTTAT
2941 AAATTCTTTT CAAGGATTCA ATAATAAATT GCCTTACGAG ATTCAGTACA TTATACTGGA
3001 GAATCTTAAT AACCATGAAC TAAAAAAAAT TTTAGATAAT ATACATTAAA AAGGTAAATA
3061 GATCATCTGT TATTATAAGC AAAGATGCTT GTTGCCAATA ATATACAACA GGTATTTGTT
3121 TTTATTTTTA ACTACATATT TGATGTTCAT TCTCTTTATA TAGTATACAC AGAAAATTCA
3181 TAATCCACTT AGAATTTCTA GTTATCTAG
```

FIG.8

FIGURE 14. A DNA sequence provided in patent #5,833,975

The original patent assignee is a pharmaceutical firm, Virogenetics Corp, that received financial support from the US government. Unfortunately, government funding for this patent cannot be traced back to a specific agency: the statement of government interest is too generic, as can be seen in Figure 15.

Most of the citations to academic work are to articles published in virology, molecular biology and immunology journals. It is worth noting that pharmaceutical and medical patents are heavily represented among patents with large shares of citations to scientific papers. Out of the top 10 patents in shares of citations to science, eight of them are either supported by the Department of Health and Human Services or are protecting health-related technologies. This reliance of medical patents on science can also be seen in the heterogeneity analysis in the top-right corner of panel C.12.

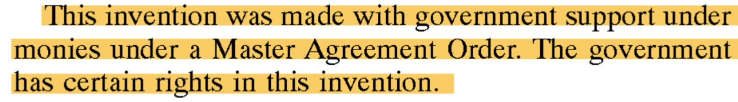

This invention was made with government support under monies under a Master Agreement Order. The government has certain rights in this invention.

FIGURE 15. Statement of government interest in patent #5,833,975

C.3.2. *Case study 2 – A random number generator before the computer era.* Patent number 4,183,088, entitled 'Random number generator', is the publicly-funded patent that predates the creation of its patent class by the longest time in my sample.[66] It was filed in 1962 by the US Navy, 37 years before being re-classified into the 'Electrical computers: arithmetic processing and calculating' USPC class upon its introduction, in 1999. Originally, it was filed under the 'Oscillators' patent class in the USPC system (class number: 331). Its subclass was 'Electrical noise or random wave generator' (78). The technology described in the patent indeed relies on a noise signal fed into a device that then combines it with another signal supplied by a pulse generator. Through a sequence of mechanical and electrical transformations of the two signals, the device provides a random sequence of ones and zeroes with a specific probability distribution to its user.

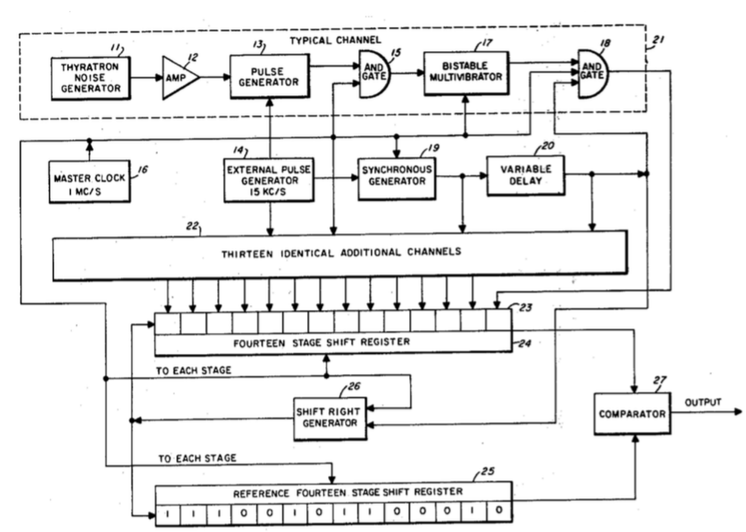


FIGURE 16. Drawing of the random number generator device of patent #4,183,088

This patent predates the computer era by several decades. The first mention of the word 'Computer' in a USPC patent class title was in 1993 in the 'Computer graphics processing and selective visual display systems' class.

[66] The sample of patents is restricted to patents filed after 1950 and to patents that are filed before their latest class is created (*i.e.* patents that are 'ahead of their time') here.

C.3.3. *Case study 3 – A shape-memory alloy with applications across many technologies.* The publicly-funded patent cited across the largest number of patent classes is patent number 5,061,914, entitled 'Shape-memory alloy micro-actuator'. It was funded by NASA but the R&D was performed by a private firm. The patent was filed in 1989 and is cited by 36 distinct patent classes.

The technology described in this patent is a type of micron-sized mechanical switch. Such minuscule switches are made of metal alloys that change shape or size when heated. They return to their original state when the temperature drops back down. This innovation is useful for creating surfaces that alternate in shape, and applications of shape-memory micro-actuator are multiple. In medicine, they are used to navigate through winding paths in the body; they change shape during surgeries. In the aerospace and automotive industries, these actuators are used to adjust components like air vents or flaps without relying on complicated mechanical systems. In consumer electronics, they can be used to protect some critical components if the device heats up above a certain temperature. NASA also uses larger scale actuators to adjust the flight performance of aircrafts and space shuttles under changing temperature conditions (NASA technology transfer program website, accessed November 24, 2023).

SHAPE-MEMORY ALLOY MICRO-ACTUATOR

This invention was made with government's support under contract NAS2.12797 awarded by NASA. The government has certain rights in this invention.

FIELD OF THE INVENTION

This invention relates generally to actuator devices. More particularly, the invention relates to an actuator device for obtaining quantitative motion of a micro-mechanical element by utilizing a shape-memory alloy actuating element, and a method of producing thin films of shape-memory material.

FIGURE 17. Statement of government interest in patent #5,061,914

Appendix D. A discussion of the linear model of innovation

The interpretation of the science-technology nexus presented in section 3 is often described as the *linear model* of innovation (Bush, 1945; Maclaurin, 1953; Nelson, 1959). It posits that intellectual progress goes from science to applied research, to development, to commercialization and to diffusion. In spite of its simplicity, the linear model has been shown to be a powerful tool to explain the interaction between fundamental research and applied innovation (Godin, 2006; Balconi *et al.*, 2010; Ahmadpoor and Jones, 2017), and most modern research takes the upstreamness of basic research *vis-à-vis* applied innovation as given (Akcigit *et al.*, 2020; Arora *et al.*, 2021a).

Appendix E. Historical SSIV – Additional results

E.1. **Summary statistics.** Table E.16 shows summary statistics on the sample of firms used in the SSIV specifications.

Variable	Mean	SD	Min	p_{10}	p_{25}	p_{50}	p_{75}	p_{90}	Max
Monetary values – million of 2020 USD									
Sales	7,844	25,028	1	71	313	1,273	5,235	16,067	498,518
Capital	4,428	16,692	0	15	70	352	2,220	9,257	375,924
Market value	8,226	27,492	0	41	170	962	4,385	16,300	702,025
R&D expenses	185	831	0	0	0	4	49	251	14,245
Counts									
Employment ('000s)	23	66	0.003	0.3	1	5	18	51	2,100
Patent count at t (flow)	47	202	0	0	0	3	19	80	4,437
Endogenous treatments and instruments									
Public spillovers	4.391	0.782	0.000	3.702	3.998	4.339	4.708	5.238	7.971
Public R&D funding	7.018	1.438	0.000	5.133	6.653	7.284	7.787	8.364	11.210
Private spillovers	0.280	0.113	0.005	0.137	0.206	0.273	0.340	0.421	1.067

States (top 5)		*Periods (top 5)*	
CA	10.5 %	2005	12.9 %
NY	8.0 %	2010	11.9 %
TX	7.8 %	2000	10.9 %
OH	7.6 %	1990	9.4 %
IL	7.4 %	1995	9.3 %

Sectors (top 5)	
367 – Electronic Components & Accessories	6.0 %
382 – Lab Apparatus & Analytical, Optical, Measuring, & Controlling Instruments	4.6 %
384 – Surgical, Medical, & Dental Instruments and Supplies	4.4 %
371 – Motor Vehicles and Motor Vehicle Equipment	3.9 %
357 – Computer & Office Equipment	3.1 %

Table E.16. Summary Statistics – SSIV sample

Notes: The unit of observation is a firm $\times$ year. Summary statistics are computed on the sample used in Table 1 for the SSIV regressions ($N = 7,631$). Monetary values are deflated using the BLS Consumer Price Index and expressed in 2020 USD.

	(1)	(2)	(3)	(4)
Productivity				
Δ_{10} ln(TFP)	.024*	.027**	.027**	.024*
	(.013)	(.013)	(.013)	(.014)
Firm sales and employment				
Δ_{10} ln(sales)	-.020*	-.015	-.017	-.015
	(.011)	(.011)	(.012)	(.010)
Δ_{10} ln(employment)	-.026**	-.021**	-.023**	-.02**
	(.011)	(.010)	(.011)	(.010)
First-stage F-stat (exp. robust)	97.34	97.40	98.14	108.14
Period FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
Sectors FE (2-digit)	✓	✓	✓	
Sectors FE (3-digit)				✓
Own R&D and patents	✓	✓	✓	✓
Private R&D spillovers		✓	✓	✓
Lagged firm controls			✓	✓
N	6,499	6,499	6,499	6,499

TABLE E.17. Historical SSIV regression results – 10-year outcomes

Notes: The unit of observation is a firm × period. Standard errors and F-stats are exposure-robust (Adão *et al.*, 2019): they are computed using the authors' `reg_ss` and `ivreg_ss` commands.
***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.

E.2. **10-year outcomes.**

E.3. **Narrative shocks.** This section describes the funding shocks I use in my robustness SSIV result. The selection of shocks is based on the historical description of R&D funding appropriations in the appendix of Fieldhouse and Mertens (2023) and my own reading of the histories of federal agencies. Table E.18 and E.19 below describes the shocks included in the instrument used for this robustness check, along with a justification for their inclusion.

NSF	
1950 1955 1960	USSR's first atomic test in 1949 + Scientific and technological competition with the USSR (ballistic missiles) + Sputnik (1957)
1980	Reagan's expansion of NSF
1990 1995 2000	Human Genome Project + 21st Century Research Fund initiative + Anthrax terrorist attacks of 2001
2010	Recovery Act

Department of Energy and Environmental Protection Agency	
1950 1955 1960	Eisenhower's 'Atoms for Peace' (advance domestic energy production, re-purpose breakthrough in fusion obtained during WWII)
1970 1975	Oil shock → more research into alternative sources of energy (motivated by energy inflation and concerns over national security)
2005	07-08 oil price shock
2010	Budget Control Act of 2011 (debt ceiling crisis)

Department of Homeland Security	
2000 2005	9/11

TABLE E.18. Shocks kept in the narrative approach (NSF, Department of Energy + Environmental Protection Agency, Department of Homeland Security)

Notes: The table shows the set of funding shocks kept in the construction of the SSIV instrument used in the 'narrative' approach robustness check. Shocks are selected based on the historical description of R&D funding across federal agencies in the appendix of Fieldhouse and Mertens (2023), and my own reading of the histories of the agencies. The right column provides a justification for the inclusion of the shock in the narrative-SSIV instrument. Justifications that are used for several consecutive five-year periods within agencies are give the same color (light gray or white).

APPENDIX F. PATENT EXAMINER REGRESSIONS – ADDITIONAL RESULTS

F.1. Sample of firms: Summary statistics.

Department of Defense	
1940	WWII
1945	WWII drawdown
1950	Korean War (1950-1953)
1955	
1960	
1965	Vietnam war (1955-1975) and drawdown (post 1975)
1970	
1975	
1980	
1985	
1990	Reagan's buildup + Russian invasion of Afghanistan + Cold War drawdown
1995	
2000	
2005	9/11 + Iraq + Afghanistan
2010	

NASA	
1955	Creation of NASA
1960	Sputnik (1957) + Apollo space program
1965	Apollo space program drawdown
1970	Loss of interest in spaceflight by Congress after the moon landing
1985	George H.W. Bush's push for NASA funding + MIR space station
1990	
2010	Budget Control Act of 2011 (debt ceiling crisis)

Department of Health and Human Services	
1970	Nixon's 'war on cancer'
1985	Reagan's push for funding during the AIDS/HIV epidemic
1990	
1995	Human Genome Project + 21st Century Research Fund initiative + Anthrax terrorist attacks of 2001
2000	
2005	Recovery Act of 2009 + Budget Control Act of 2011 (debt ceiling crisis)
2010	

TABLE E.19. Shocks kept in the narrative approach (DoD, NASA, Department of HHS)

Notes: The table shows the set of funding shocks kept in the construction of the SSIV instrument used in the 'narrative' approach robustness check. Shocks are selected based on the historical description of R&D funding across federal agencies in the appendix of Fieldhouse and Mertens (2023), and my own reading of the histories of the agencies. The right column provides a justification for the inclusion of the shock in the narrative-SSIV instrument. Justifications that are used for several consecutive five-year periods within agencies are give the same color (light gray or white).

Variable	Mean	SD	Min	p_{10}	p_{25}	p_{50}	p_{75}	p_{90}	Max
Monetary values – million of 2020 USD									
Sales	7,009	24,259	0	32	166	976	4,039	14,795	498,518
Capital	4,072	14,726	0	5	35	254	1,877	8,940	294,387
Market value	9,113	30,348	0	28	148	924	4,425	18,157	702,025
R&D expenses	162	852	-0	0	0	0	19	142	13,045
Counts									
Employment ('000s)	18	63	0	0	1	3	12	43	2,100
Patent count	33	201	0	0	0	0	3	30	4,437
Endogenous treatments and instruments									
Private spillovers	0.137	0.164	0.000	0.000	0.000	0.000	0.299	0.356	0.722
Public spillovers	1.057	1.289	0.000	0.000	0.000	0.000	2.323	2.684	6.447
Private leniency	0.000	0.001	-0.013	-0.001	0.000	0.000	0.000	0.001	0.014
Public leniency	-0.003	0.009	-0.162	-0.011	-0.005	0.000	0.000	0.000	0.082

States (top 5)		*Periods*	
CA	11.0 %	2005	34.3 %
TX	9.7 %	2010	33.3 %
NY	7.5 %	2015	32.4 %
OH	4.6 %		
MA	4.5 %		

Sectors (top 5)	
737 – Computer Programming, Data Processing, & other Computer Services	6.7 %
367 – Electronic Components & Accessories	4.6 %
283 – Drugs	4.0 %
491 – Electric Services	3.8 %
384 – Surgical, Medical, & Dental Instruments and Supplies	3.6 %

TABLE F.20. Summary Statistics – Patent examiner IV sample

Notes: The unit of observation is a firm × year. Summary statistics are computed on the sample used in Table 4 for the patent examiner IV regressions ($N = 2,118$). Monetary values are deflated using the BLS Consumer Price Index and expressed in 2020 USD.

APPENDIX G. PROOFS AND DERIVATIONS

G.1. **Summary of the notation used in the model.** Table G.22 summarizes the notation used in the theory section.

	(1)	(2)	(3)
Application is publicly-funded	0.0001	0.0004	0.0005
	(0.0010)	(0.0009)	(0.0009)
Art unit FE	✓	✓	✓
Art unit × year FE		✓	✓
Patent count of applicant in current year			✓
Mean dep. var.	0.73	0.73	0.73
R^2	0.552	0.620	0.6120
N	681,023	681,023	681,023

TABLE F.21. Are government applicants favored by USPTO examiners?

Notes: The unit of observation is a patent application × year. The table shows the results of a regression of examiner leniency on a dummy variable equal to 1 if the application is funded by public R&D. The years in the sample are those used in the patent examiner regressions *i.e.* 2001, 2005 and 2010. ***, **, and * indicate two-sided significance at the 1, 5 and 10% levels, respectively.

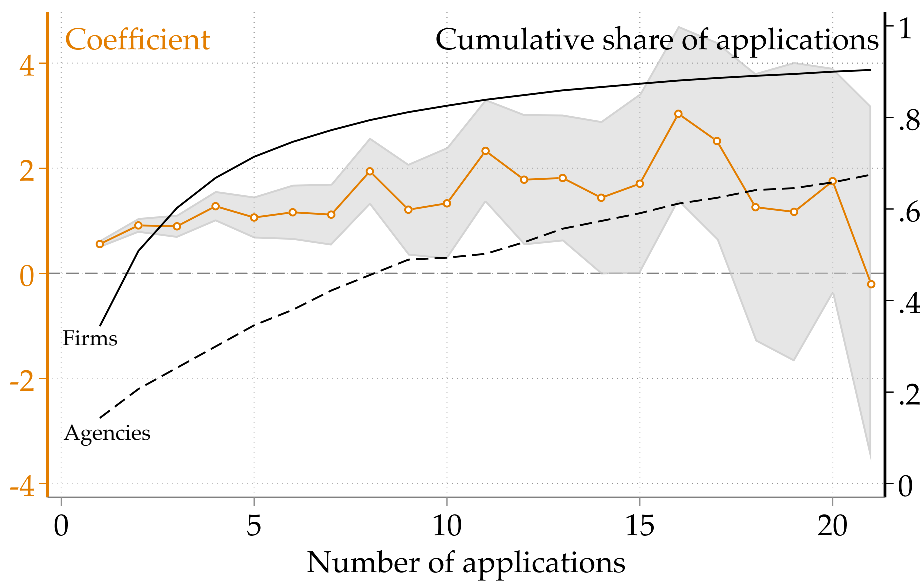
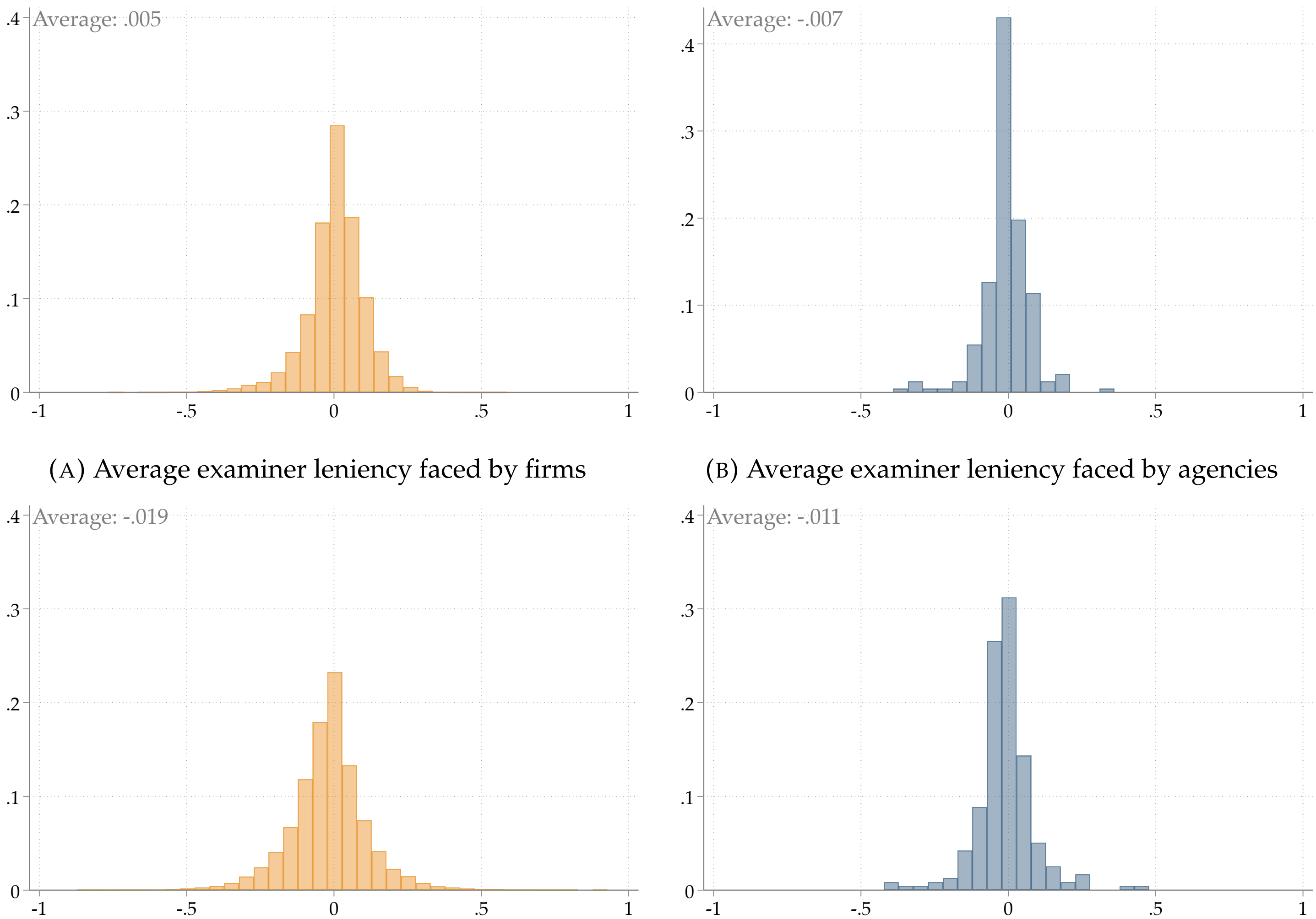


FIGURE 18. Diminishing strength of the first stage for large entities

Notes: The graph shows how the strength of the relationship between the growth of patents of a firm and the patent examiner instrument (changes in average leniency) evolves as entities submit more and more applications. The unit of analysis is an entity (firm or agency) j in a five-year period t. The orange line and shaded area show the coefficients and 95% confidence intervals coming from a regression of Δp_{jt} on $\Delta\overline{l_{jt}}$. This is the variation underlying the patent examiner IV strategy. In my regressions reported in the main text, Δp_{jt} and $\Delta\overline{l_{jt}}$ are then aggregated across *receiving* firms (indexed by i in the main text). The solid and dashed lines show the cumulative distributions of entities × year across their numbers of applications, for firms and public entities respectively. The distribution of agencies first-order stochastically dominates that of firms because firms tend to file fewer patents than agencies.



(A) Average examiner leniency faced by firms

(B) Average examiner leniency faced by agencies

(C) 5-year difference in firm leniency

(D) 5-year difference in agency leniency

FIGURE 19. Histograms describing the patent examiner variation

Notes: The histograms show the distributions of average leniencies faced by firms and agencies (panels 19a and 19a respectively) and the 5-year differences in average leniencies faced by firms and agencies (panels 19c and 19c respectively). By construction, the average leniency is centered around 0; it is the firm- or agency-level average of residuals of a regression of an examiner leniency on art unit fixed effects.

G.2. **Key model equations.**

Government	
τ	Tax rate on firms' profits
Production	
w	Production wage rate
α	Drift of firms' productivity
ν	SD of firms' productivity
σ	SD of firms' normalized profits
g	Growth rate of the aggregate economy
e	Private research effort
ϕ_0	Returns to (applied) R&D effort
ρ	Discount rate of firm owners
ζ	Pareto tail exponent
ξ	Power law inequality ($\xi = 1/\zeta$)
A	Aggregate productivity index
δ	Endogenous exit rate ('creative destruction')
$\bar{\delta}$	Exogenous (baseline) exit rate
Innovation and spillovers	
β_g and β_i	Indicators of the type of research funded by the government and firms
ε	Elasticity of productivity to applied spillovers
γ	Elasticity of productivity to basic spillovers
Γ	Innovation step size
Ψ	Aggregate growth component
w_g	Research wage, publicly-funded researchers
w_p	Research wage, privately-funded researchers
Λ	Private=public wage premium
λ	Arrival rate of ideas
χ	Share of ideas from spillovers successfully turned into businesses
Households	
L_t	Population at t
θ	Substitution parameter of intermediate varieties (elasticity of subs. = $1/(1-\theta)$)

TABLE G.22. Notation used in the model

G.3. Proof of lemma 1.

Proof. **Labor demand and intermediate output** The final sector's problem is:

$$\max_{y_i} \left(\int_0^1 y_i^\theta \mathrm{d}i \right)^{\frac{1}{\theta}} - \int_0^1 p_i y_i \mathrm{d}i \qquad \forall i \in [0,1] \tag{24}$$

First order conditions with respect to y_i give $\theta y_i^{\theta-1} \frac{1}{\theta} \left(\int_0^1 y_i^\theta \mathrm{d}i \right)^{\frac{1}{\theta}-1} - p_i = 0$ and the inverse demand for y_i is thus:

Description	Equation
Optimization	
Intermediate output choice	$y_i = Y\,(a_i/A)^{\frac{1}{1-\theta}}$
Labor choice	$l_i = (a_i^\theta/A)^{\frac{1}{1-\theta}}\, Y/\Psi$
Research effort	$e = 1-\tau-\frac{1-\theta}{\theta}\frac{\rho+\delta+\bar{\delta}}{\phi_0}$
Choices of type of research	$\beta_g = 1$ and $\beta_i = 0 \quad \forall i$
Law of motion of productivity	$\mathrm{d}a_{it}/a_{it} = e\phi_0 \mathrm{d}t + \nu \mathrm{d}B_t$
Resource constraint	
Allocation of research personnel	$R_g = \frac{R}{e/\Lambda\tau+1}$ and $R_p = \frac{R}{\Lambda\tau/e+1}$
Aggregation and equilibrium objects	
Labour market clearing condition	$L := \int_0^1 l_i \mathrm{d}i$
Definition of aggregate output	$Y := \left(\int_0^1 y_i^\theta \mathrm{d}i\right)^{\frac{1}{\theta}}$
Effect of spillovers on the economy	
Definition of spillovers	$\dot{n}_t := \ln(\lambda R_g)^\gamma(\lambda R_p)^\epsilon$
Definition of creative destruction	$\delta := \chi\dot{n}_t$

TABLE G.23. Key model equations

Notes: The endogenous variables of interest are $Y, y_i, a_i, L, l_i, e, R_p, R_g, \dot{n}, \delta, \beta_g, \beta_i$. Time subscripts are omitted when it does not cause confusion.

$$p_i = \left(\frac{Y}{y_i}\right)^{1-\theta} \tag{25}$$

Plugging (25) into the objective function of monopolist i, replacing l_i by y_i/z_i and taking first order conditions with respect to y_i, I obtain the profit-maximizing output level for a firm with productivity z_i:

$$y_i^* = Y\left(\frac{\theta}{w}\right)^{\frac{1}{1-\theta}} z_i^{\frac{1}{1-\theta}} \tag{26}$$

and because $y_i^* = z_i l_i^*$, labor demand is:

$$l_i^* = Y\left(\frac{\theta}{w}\right)^{\frac{1}{1-\theta}} z_i^{\frac{\theta}{1-\theta}} \tag{27}$$

Equilibrium wage w and aggregate output Y. The equilibrium wage w is obtained by plugging (26) into the definition of final output (9), which gives:

$$w = \theta A \Psi \tag{28}$$

where $A = \left(\int_0^1 a_i^{\frac{\theta}{1-\theta}} \mathrm{d}i\right)^{\frac{1-\theta}{\theta}}$ is the (idiosyncratic) productivity index of the economy.[67] The value of Y in (26) and (27) can be obtained by plugging the expression for l_i^* (27) into the labor market clearing condition $\int_0^1 l_i \mathrm{d}i = L$ and using the expression for the wage rate (28). I obtain the following expression for the equilibrium value of aggregate output:

$$Y = LA\Psi \tag{29}$$

This proves part 3 of lemma 1. Then, using (28), intermediate output and labor demand can be written more simply as:

$$y_i^* = Y\left(\frac{a_i}{A}\right)^{\frac{1}{1-\theta}} \qquad \text{and} \qquad l_i^* = \frac{Y}{\Psi}\left(\frac{a_i^{\theta}}{A}\right)^{\frac{1}{1-\theta}}$$

Which proves part 1 of lemma 1.

Firm profits π_i^* and wage bill wl_i^*. Firm profits are, by definition,

$$\pi_i^* = p_i y_i^* - wl_i^* \tag{30}$$

and their value as a function of real variables is given by replacing p_i by (25), l_i^* by (27) and w by its equilibrium value (28). Then, replacing y_i by (26) and $\frac{\theta}{w}$ by $\frac{1}{A\Psi}$ (from (28)) gives a simple expression of profits, which are equal to a $1-\theta$ share of revenues:

$$\pi_i^* = Y\left(\frac{a_i}{A}\right)^{\frac{\theta}{1-\theta}}(1-\theta) \tag{31}$$

Conversely, the wage bill of firm i is a θ share of its revenues. Its expression is obtained by plugging the equilibrium value of l_i^* in (27) into wl_i^* and then replacing w by its expression given by equation (28)

$$wl_i^* = Y\left(\frac{a_i}{A}\right)^{\frac{\theta}{1-\theta}}\theta \tag{32}$$

This proves part 2 of lemma 1 and thus completes the proof. □

G.4. A useful lemma regarding the law of motion of profits.

[67] The productivity index of the economy is the power mean of firms' idiosyncratic productivities, where the power $\frac{\theta}{1-\theta}$ increases in the substitutability of varieties. By properties of power means, A is increasing in subsitutability: the intuition is that when substitution between varieties becomes easier, the final good producer buys more from the highest-productivity firm (exclusively from it when $\theta = 1$ *i.e.* in the case where varieties are perfect substitutes).

Lemma 3 (Law of motion of profits). *On a balanced growth path, if productivity evolves as* (11)*, then profits evolve as*

$$\frac{\mathrm{d}\pi_{it}}{\pi_{it}} = \mu(e_{it}, \beta_{it})\mathrm{d}t + \sigma \mathrm{d}B_t \tag{33}$$

with drift $\mu(e_{it}, \beta_{it}) := \frac{\theta}{1-\theta}(\alpha(e_{it}, \beta_{it}) + g_\Psi)$ *and standard deviation rate* $\sigma := \frac{\theta}{1-\theta}\nu$

G.5. **Proof of lemma 3.**

Proof. From lemma 1, I know that profits are $\pi^*_{it} = Y_t \left(\frac{a_{it}}{A_t}\right)^{\frac{\theta}{1-\theta}}(1-\theta)$. Taking logs and time derivatives, I get that the long-run growth rate of a firm's profits is equal to

$$g_\pi = g_Y + \frac{\theta}{1-\theta} g_a$$

where g_x stands for the instantaneous growth rate of variable x. A_t is constant because, on a BGP, the distribution of idiosyncratic firm productivities is stationary. Therefore, A_t does not contribute to the growth of profits.

To find the value of g_Y, I rely on the expression of Y_t provided by lemma 1 which has shown that $Y_t = L_t A_t \Psi_t$, so $g_Y = g_\Psi$ on a BGP where there is no population growth.

A firm's idiosyncratic productivity drift is given by $g_a = \alpha(e_{it}, \beta_{it})$. Therefore,

$$g_\pi = g_\Psi + \frac{\theta}{1-\theta}\alpha(e_{it}, \beta_{it})$$

Turning to the standard deviation of normalized profits, I note that its value depends on the only stochastic term in the expression of π_{it}: a_{it}. Noting that, on a BGP, $a_{it} = a_{i,0}e^{\alpha(e,\beta)t+\nu B_t}$, I get that $a_{it}^{\frac{\theta}{1-\theta}} = \left(a_{i,0}e^{\alpha(e,S)t+\nu B_t}\right)^{\frac{\theta}{1-\theta}}$. Therefore the standard deviation rate of $a_{it}^{\frac{\theta}{1-\theta}}$ is $\frac{\theta}{1-\theta}\nu$. Consequently, $\frac{\theta}{1-\theta}\nu$ is the standard deviation rate of profits.

□

G.6. Proof of proposition 1. The proof of this proposition proceeds in four steps. I start by showing that the government invests exclusively in basic R&D because this maximizes the arrival rate of breakthrough innovations. Then turning to firms, I provide a closed-form expression of the value function of firms that is then used to show that firms only invest in applied research. Finally, I shows that the level of research effort exerted by firms is a constant share of profits for all firms and that it is decreasing in the tax rate τ at a given level of spillovers.

Proof. I start by showing that $R_g = R_{gb}$, that is, all researchers paid by the government are doing basic research.

Given an exogenous tax rate τ, the government raises revenues $\tau\Pi$ where Π is the aggregate flow of profits in the economy. The government seeks to maximize the arrival rate of breakthroughs which is the sum of the flows of breakthroughs from basic and applied research: $\lambda_1 R_1 +$

$\lambda_0 R_0$. Because the breakthrough Poisson rate per researcher is higher for basic research than for applied research ($\lambda_1 > \lambda_0$) and the wage of researchers is common across basic and applied researchers, the allocation of researchers that maximizes breakthrough flow is, trivially, a corner solution where all government-funded researchers are doing basic research.

This proof follows the argument in the proof of proposition 1 in Jones and Kim (2018). The HJB reads

$$(\rho + \delta + \bar{\delta})v(a,t) = \max_{e,\beta} \quad \ln(\Psi a^{\frac{\theta}{1-\theta}}) + \ln(1 - e - \tau) + \alpha(e,\beta) a v_a(a,t) + \frac{\nu^2}{2} a^2 v_{aa}(a,t) + v_t(\pi, t)$$

Taking first order conditions of the HJB with respect to e gives

$$\frac{1}{1 - e - \tau} = \phi(\beta) a v_a(a,t) \tag{34}$$

I guess and verify that the value function takes the form $v(a,t) = \alpha_0 + \alpha_1 t + \alpha_2 \ln(a)$. Using this functional form for $v(a,t)$, (34) becomes

$$\frac{1}{1 - e - \tau} = \phi(\beta)\alpha_2 \tag{35}$$

Using (35) and the guess for the functional form of the HJB gives

$$(\rho + \delta + \bar{\delta})(\alpha_0 + \alpha_1 t + \alpha_2 \ln(a)) = \frac{\theta}{1-\theta}\ln(a) + \ln(Y(1-\theta)A^{\frac{\theta}{\theta-1}}) + \ln(1 - e - \tau) + e\phi(\beta)\alpha_2 - \frac{\nu^2}{2}\alpha_2 + \alpha_1$$

Equating coefficients on $\ln(a)$ gives: $\alpha_2 = \dfrac{\theta}{(1-\theta)(\rho + \delta + \bar{\delta})}$. Plugging this value of α_2 into (35) gives the optimal R&D effort level

$$e^* = 1 - \tau - \frac{1-\theta}{\theta}\frac{\rho + \delta + \bar{\delta}}{\phi(\beta)} \tag{36}$$

This proves the third point of proposition 1.

To show that the HJB equation is linear in t, as posited by the conjecture, I first note that the only term other than $\ln(a)$ that depends on time is $\ln(Y)$. As shown in lemma 1, $Y = LA\Psi$ with $\Psi = \Gamma^{n_t}$. In a balanced-growth path equilibrium, the flow rate of ideas $\dot{n}_t$ is constant, so n_t is linear in t. This proves that $\ln(Y)$ is linear in t.

For completeness, the value function of a firm with productivity a is $v(a,t) = \alpha_0 + \alpha_1 t + \alpha_2 \ln(a)$ with

$$\begin{aligned}
\alpha_0 &= C\ln\left(L(1-\theta)A^{\frac{\theta}{\theta-1}+1}(1-e^*-\tau)\right) + C^2\left(e^*\phi(\beta) - \frac{\nu^2}{2}\right)\frac{\theta}{1-\theta} + C\alpha_1 \\
\alpha_1 &= C\ln(\Gamma)\ln((\lambda R_p)^{\varepsilon}(\lambda R_b)^{\gamma}) \\
\alpha_2 &= C\frac{\theta}{1-\theta} \\
\text{with} \quad C &= \frac{1}{\rho+\delta+\overline{\delta}}
\end{aligned} \tag{37}$$

This step completes the derivation of the value function.

To prove that firms only invest in applied research, one notes that the value function is strictly increasing in $\phi(\beta)$ at every level of research effort. Because $\phi_0 > \phi_1$, firm owners only invest in applied research. This proves part (2) of the proposition.

□

G.7. **Proof of lemma 2.**

Proof. To find the stationary distribution of firms satisfying the KFE (18), guess that f takes the form $f(a) = Ca^{-\zeta-1}$, where C is a positive constant. Insert this candidate solution in (18) and get

$$0 = -\overline{\delta}Ca^{-\zeta-1} - \alpha\partial_a[Ca^{-\zeta}] + \frac{\nu^2}{2}\partial_{aa}[Ca^{-\zeta+1}] \tag{38}$$

$$0 = -\overline{\delta}Ca^{-\zeta-1} + \alpha\zeta Ca^{-\zeta-1} - \frac{\nu^2}{2}(1-\zeta)\zeta Ca^{-\zeta-1} \tag{39}$$

$$0 = -\overline{\delta} + \alpha\zeta - \frac{\nu^2}{2}(1-\zeta)\zeta \tag{40}$$

where α is shorthand for $\alpha(e^*, \beta^*)$.

This equation admits two solutions for ζ which are

$$\zeta^{\pm} = -\frac{\alpha}{\nu^2} + \frac{1}{2} \pm \sqrt{\left(\frac{\alpha}{\nu^2} - \frac{1}{2}\right)^2 + \frac{2\delta}{\nu^2}}$$

The positive root is the only one consistent with a CDF that is a convergent integral.

Furthermore, the constant C is given by the requirement that the mass of firms integrates to 1.

$$\int_{a_0}^{\infty} C a^{-\zeta-1} \mathrm{d}a = 1$$

$$C \left[\frac{a^{-\zeta}}{-\zeta} \right]_{a_0}^{\infty} = 1$$

$$C \left(\lim_{z \to \infty} \frac{z^{-\zeta}}{-\zeta} + \frac{a_0^{-\zeta}}{\zeta} \right) = 1$$

$$C = \zeta a_0{}^{\zeta}$$

□

G.8. **Proof of proposition 2.**

Proof. On a BGP, the rate of creative destruction is $\delta = \chi \dot{n} = \ln((\lambda R_g)^{\gamma} (\lambda R_p)^{\varepsilon})$. Replacing R_g and R_p by the expressions in (22), taking derivatives with respect to τ and noting that $\partial e^* / \partial \tau = -1$ from (15), I obtain:

$$\frac{\partial \delta}{\partial \tau} = \underbrace{R \frac{\gamma}{\Lambda} \frac{1}{e^*/\tau\Lambda + 1} \frac{\tau + e^*}{\tau^2}}_{\text{marginal gain from public R\&D}} - \underbrace{R \varepsilon \Lambda \frac{1}{\tau\Lambda / e^* + 1} \frac{\tau + e^*}{e^{*2}}}_{\text{marginal loss from private R\&D}}$$

The first term in the difference capture the (positive) impact of raising the tax rate on creative destruction through the contribution of publicly-funded research. The second term captures the declining contribution of privately-funded research to creative destruction when the tax rate increases.

Setting $\frac{\partial \delta}{\partial \tau}$ equal to 0 and solving for τ gives

$$\tau^* = \frac{\gamma e^*}{\varepsilon \Lambda}$$

For values of τ in $[0, \tau*)$, $\frac{\partial \delta}{\partial \tau}$ is positive and the rate of creative destruction is increasing in the tax rate. For values of τ in $(\tau*, 0]$, $\frac{\partial \delta}{\partial \tau}$ is negative. This shows that δ is inverted-U-shaped in the tax rate.

From (19), one gets that ζ is increasing in δ and thus Pareto inequality η is decreasing in δ. Inequality is minimized when δ is highest *i.e.* when $\tau^* = \frac{\gamma e^*}{\varepsilon \Lambda}$. Plugging the value of τ^* into (22) gives $R_g{}^*$. This proves (1) and the inequality part of (3).

To show that the growth rate is inverted-U-shaped in the tax rate, I note that $\frac{\partial g}{\partial \tau} = \frac{1-\theta}{\theta} \ln(\gamma) \frac{\partial \delta}{\partial \tau}$ because $\delta = \dot{n}_t$. Hence the comparative statics of g with respect to τ are the same as those for δ.

Therefore g is growing in $\tau \in [0, \tau*)$, decreasing in $\tau \in (\tau*, 0]$ and maximized at τ^*. This proves (2) as well as the growth part of (3) and thus completes the proof.

$\square$

Appendix H. Calibration

H.1. Data. Data is annual. The historical TFP series come from Bergeaud *et al.* (2016) and is calculated assuming a Cobb-Douglas aggregate production function with capital and labor inputs.[68] Data on inequality between firms come from Kwon *et al.* (2022), who digitized archival records from the US Internal Revenue Service. I use their series on firm assets to measure firm inequality as it is continuous over the period of study (unlike their series on net income and receipts). I then calculate the empirical Pareto tail exponent ζ_{data} by using an insight from Chen (2022): with the share of assets s_x of the top x% firms, one can estimate the tail exponent as:

$$\zeta_{\text{data}} = \left(1 - \frac{\ln(s_{x_1}/s_{x_2})}{\ln(x_1/x_2)}\right)^{-1}$$

In my application, I use $x_1 = 10$ and $x_2 = 1$ so that inequality between firms is a function of inequality between the top 10 and the top 1% of firms, by assets.

The tax rate τ (the main exogenous parameter of interest) is set to be a direct function of public R&D spending: it evolves in concert with public R&D as a share of total R&D. I set the value of τ equal to the effective corporate tax rate in the US in 1947, when the data is first available[69]. The value of τ in the following years is then given by

$$\tau = \text{share of public R\&D in total R\&D} \times \frac{\text{effective corporate tax rate at } t = 0}{\text{share of public R\&D in total R\&D at } t = 0}$$

The tax rate calculated in this way closely follows the effective tax rate, as can be seen in Figure 20.

London School of Economics, department of Economics

[68]Formally, $TFP = \frac{Y}{K^{\alpha}L^{1-\alpha}}$. Aggregate capital is the sum of 'equipment' and 'buildings', from the National Accounts (BEA). Aggregate labor is the total number of hours worked (from various academic sources).

[69]The effective corporate tax rate is $\frac{\text{aggregate profits before tax - aggregate profits after tax}}{\text{aggregate profits before tax}}$. The effective tax rate will be lower than the statutory tax rate if deductions, tax credits (from previous losses or from R&D credits for instance) and tax avoidance schemes lower the tax burden of firms. It is a more representative measure of the tax burden faced by firms. Data on total corporate profits before and after tax come from the BEA series 'Corporate profits before tax (without IVA and CCAdj)' and 'Corporate Profits After Tax (without IVA and CCAdj)', respectively.

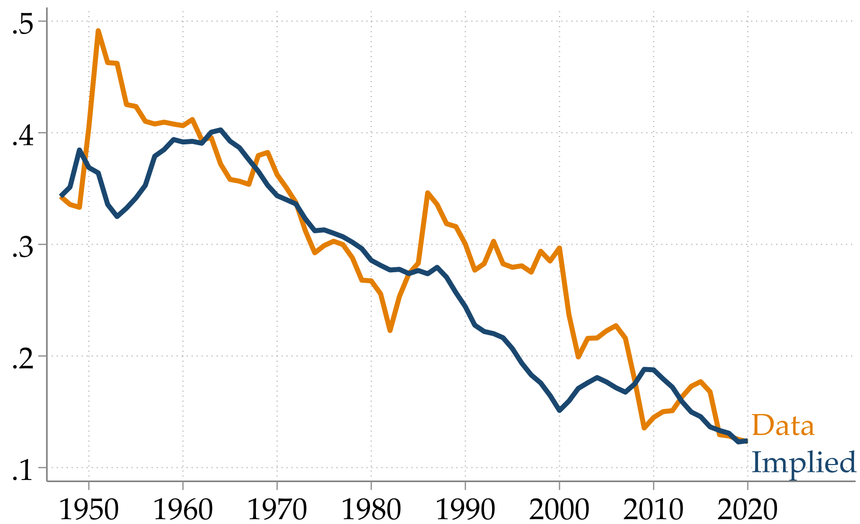


FIGURE 20. Effective tax rate in the US (orange) and tax rate used in the model (blue)

Notes: The effective corporate tax rate is $\frac{\text{aggregate profits before tax - aggregate profits after tax}}{\text{aggregate profits before tax}}$. The effective tax rate will be lower than the statutory tax rate if deductions, tax credits (from previous losses or from R&D credits for instance) and tax avoidance schemes lower the tax burden of firms. It is a more representative measure of the tax burden faced by firms. Data on total corporate profits before and after tax come from the BEA series 'Corporate profits before tax (without IVA and CCAdj)' and 'Corporate Profits After Tax (without IVA and CCAdj)', respectively.