Ph. D. Dissertation in Engineering

Essays on scientific knowledge and inventor mobility

February 2022

Graduate School of Seoul National University
Technology Management, Economics, and Policy Program
Seungryul Ryan Shin

Essays on scientific knowledge and inventor mobility

지도교수 황준석

이 논문을 공학박사학위 논문으로 제출함
2022 년 2 월

서울대학교 대학원
협동과정 기술경영경제정책 전공
신 승 렬

신승렬의 공학박사학위 논문을 인준함
2022 년 2 월

위 원 장 **Yeonbae Kim** (인)

부위원장 **Junseok Hwang** (인)

위 원 **Keun Lee** (인)

위 원 **Benjamin Balsmeier** (인)

위 원 **Lee Fleming** (인)

Abstract

Essays on scientific knowledge and inventor mobility

Seungryul Ryan Shin

Technology Management, Economics, and Policy Program

College of Engineering

Seoul National University

Facilitating technological innovation in various fields and stimulating economic growth based on innovation is vital today. Both scientific knowledge and human capital are key components in seeking technological innovation. This dissertation studies the mechanisms of innovation from the aspects of scientific knowledge diffusion and inventor mobility. Each chapter here investigates the diffusion of government science, causes, and implications of inventor mobility. This dissertation develops and applies data analytics and causal inference techniques to conduct empirical analyses for each chapter. This dissertation also expands the academic discussion on innovation and suggests policy and managerial implications conducive to innovation and economic growth.

The first chapter of the dissertation examines the role of a patent filed by government scientists in the dissemination of scientific discoveries in government laboratories. It collects a data sample on scientific discoveries originating from US federally funded

R&D centers and adopt a difference-in-differences approach to compare similar scientific discoveries. The results confirm that while a patent filed by government scientists decreases the rate of follow-on patents in a technological area overlapping with those of the focal patent, it increases the rate of follow-on patents in non-overlapping technological areas. Increase in follow-on inventions is attributable to risk-taking inventions, that is, inventions involving a high chance of resulting in either impactful or failure patents rather than incremental inventions. It is also characterized by inventions with a high level of originality. The results also show that inventors in distant locations in terms of geographical and technological proximity are the most affected by patents filed by government scientists. This patent effect is pronounced when government scientists involved in the focal discovery have fewer social connections and when the scientific field is less familiar in the industry.

The results of the first chapter help policymakers design policies on the patenting activities of government scientists by providing relevant empirical evidence. The results also suggest policymakers to leverage patents and implement other means to increase interaction between government scientists and industrial laboratories. Furthermore, the chapter contributes to building a complementary structure between government laboratories and industrial firms by suggesting important policy implications conducive to the diffusion of scientific discoveries in government laboratories responding to industrial landscape changes in the current era.

The second chapter of the dissertation uncovers the determinants of inventors' mobility choices. It demonstrates that an inventor's emigration from a location is negatively associated with the historical share of the same surname in a given location. This surname effect on inventors' geographical mobility is valid even after controlling for inventor individuals' length of invention experience, productivity, quality; previous mobility pattern, active technological fields; and current location characteristics. Additionally, the chapter uncovers specific conditions wherein the surname effect is moderated and wherein the surname effect loses its significance. The chapter also examines heterogeneity effects by gender, showing that surname effect differs by gender.

The results of the second chapter provide empirical evidence that inventors' geographical mobility may be predicted by historical surname distribution and how other factors influence this relationship. These findings provide empirical evidence and implications that help attract and retain inventors. Particularly, by demonstrating that the surname effect is less susceptible to other individual-level invention-related characteristics, the results increase the generalizability of using historical surname distribution to track inventors' geographical mobility.

The third chapter of the dissertation estimates the impact of inventor inflow on the formation and success of local venture-backed startups and on local shifts in venture capital investments. Moreover, it strengthens causality with a shift-share instrument based on historic spatial distribution of surnames. The results show that the arrival of inventors increases the number of venture-backed startups, but only in the same technological fields of newly arrived inventors. Inventor arrivals also increase the number of successful

startups while reducing failed ones, demonstrating that inventors improve startup quality. Furthermore, inventor inflow incurs venture capital investments to reallocate from low-tech unsuccessful startups to high-tech successful startups. This shows an increase in the efficiency of venture capital investments.

The results of the third chapter provide empirical evidence on the role of inventors in venture-backed startup activities and investment shifts. This suggests implications for policymakers who aim to foster an entrepreneurial ecosystem that they may improve the quantity and quality of startups as well as efficiency of investments by attracting inventors. Additionally, the shift-share instrument developed in this chapter would be useful in estimating the effect of human capital mobility on regional outcomes in future research.

In conclusion, this dissertation enhances the understanding of the effects of scientific knowledge and inventor mobility on innovation and economic growth. It provides several policy implications through theoretical discussions and empirical analyses centered on innovation. It also expands and contributes to a number of research streams that each chapter draws upon.

Keywords: Scientific knowledge, Government laboratory, Knowledge diffusion, Inventor mobility, Innovation, Entrepreneurship
Student Number: 2017-31119

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Backgrounds

Innovation scholars focus on scientific knowledge and human capital to enhance the understanding of technological innovation. Scientific knowledge helps us understand fundamental principles and propel further applications to technological inventions (Nelson, 1959; Kline & Rosenberg, 1986; Fleming & Sorenson, 2004). Scientific human capitals, such as scientists, inventors, or engineers, apply scientific and engineering principles to resolve technical problems, increasing useful and novel innovation (Arrow & Capron, 1959; Rosenberg & Nelson, 1994; Arts & Fleming, 2018). As technological innovation becomes key in economic growth in modern society (Lee, 2013), scholars strive to uncover the antecedents, mechanisms, and impact of innovation from the aspects of scientific knowledge and human capital, providing relevant implications for policymakers and firm managers.

Prior literature on scientific knowledge emphasizes science as a key driver of technological inventions. Fleming and Sorenson (2004) suggest that science enables effective management of R&D processes and leads to novel and useful combination of knowledge. Ahmadpoor and Jones (2017) and Poege et al. (2019) provide empirical evidence that the majority of patented inventions are based on scientific discoveries, and a patented invention is more likely to be impactful when it is closely linked to a scientific paper. They also show that the value of patented inventions is closely related to the quality of the scientific papers cited. Using data on corporate scientific research, Arora et al. (2021) provide empirical evidence on how a firm's internal scientific research contributes to its downstream inventions and, at the same time, results in spillovers to rival firms.

Scholars also seek to uncover the antecedents of scientific research. For instance, scholars examine how public investment influences scientific research (Li et al., 2017) and further leads to technological innovation (Fleming et al., 2019). Myers (2020) examines the elasticity of scientists in changing the direction of their research work, providing an estimate of the switching costs of science. Agrawal et al. (2018) investigate the change in collaboration structure of research fields in response to an influx of scientific knowledge in the field. Azoulay et al. (2019) demonstrate the evolution and direction changes in scientific fields following the loss of a luminary of the field.

Another research stream investigates scientific human capital that shapes innovation and contributes to regional economic growth. As Arrow (1962) suggests that the mobility of human capital involves the transfer of knowledge and information, scholars pay particular attention to the mobility of scientific human capital. Determinants of mobility of inventors are investigated at the individual (e.g., Palomeras and Melero, 2010; Ganco, 2013), organizational (e.g., Singh & Agrawal, 2011; Agarwal et al., 2009; Cheyre et al., 2015), and institutional levels (e.g., Marx et al. 2009; Hombert & Matray, 2017; Starr et al., 2018; Melero et al., 2020), suggesting managerial and policy implications for scientific human capital. Furthermore, scholars also examine the implications of the mobility of scientific human capital. Prior studies demonstrate how inventor mobility increases interfirm knowledge flow (e.g., Rosenkopf & Almeida, 2003), technology-oriented collaboration (e.g., Wagner & Goossen, 2018), and innovation performance (e.g., Chemmanur et al., 2019) at the organizational level. The implication of inventor mobility in productivity is also examined at the individual (e.g., Hoisl, 2007) and regional levels (e.g., Cappelli et al., 2019).

Extending prior literature on innovation, the dissertation uncovers the mechanism of dissemination of scientific research, especially those focusing on government science. It also examines the determinants and implications of the geographical mobility of inventors.

1.2 Research objectives

The dissertation aims to extend prior literature on innovation and enhance the understanding of innovation and consequent economic growth. To this end, this dissertation conducts three different studies, each of which investigates innovation from the aspects of scientific knowledge and human capital. This approach allows for a multifaceted analysis and discussion of innovation.

The first study (Chapter two) aims to understand the mechanism of dissemination of scientific knowledge. Among various sources of scientific knowledge, it focuses on government laboratories for the following reasons: 1) a significant source of scientific knowledge requiring a long-term and persistent research, 2) idiosyncratic knowledge accumulated within the laboratory, and 3) inadequate information on the technological potential available to outside researchers. A detailed examination of the mechanism of scientific knowledge dissemination is necessary to facilitate dissemination of scientific knowledge from such organizations. This study focuses on the role of patents filed by government scientists in the dissemination of government science. Adopting theoretical discussion in research streams on patent protection and relationship between science and technology, this study theorizes the mechanism by which government science is disseminated and adopted by follow-on inventions. This study also provides empirical evidence supporting the theoretical explanations.

The second study (Chapter three) aims to identify the determinants of inventors' geographical mobility. This is important as attracting incoming inventors and retaining existing inventors contributing to regional entrepreneurship (as the third study finds), which, in turn, leads to the growth of the regional economy. This study examines the effect of historical surname distribution on inventors' mobility decisions. It further investigates how historical surname distribution interacts with various other factors. This study provides empirical evidence on the determinants of mobility.

The third study (Chapter four) aims to understand how skilled human capital contributes to regional entrepreneurial ecosystems. It also aims to develop a shift-share instrument that helps isolate and estimate the causal effect of human capital on various regional outcomes. Among the various types of human capital, this study focuses on inventors with technological expertise and technological developments. For outcome variables, the study adopts the quantity and quality of startups founded and venture capital investments in high and non-high technology sectors. This study provides precise estimates of the regional effect of inventors and discusses regional entrepreneurship.

1.3 Research outline

Chapter two, three, and four provide theoretical or empirical analyses on the research questions that the dissertation addresses. These chapters study innovation from the angles of scientific knowledge and skilled human capital. Figure 1-1 describes the overall research models and how these chapters contribute to innovation literature. Table 1-1 provides a detailed overview of each chapter.

Chapter two examines how patents filed by government scientists affect follow-on inventions based on scientific discoveries in government science. It begins by reviewing relevant research streams—studies on government laboratories, patent protection, and the relationship between science and technology. It also explains the background of policy changes for US government laboratories. Then, it theoretically discusses how patent filing by government scientists influences the dissemination of scientific knowledge and follow-on inventions. Chapter three performs an empirical analysis using science data from US federally funded R&D centers (FFRDCs). It examines how patenting influences the direction and characteristics of follow-on inventions, how effect of patent filing depends on proximity of knowledge adopters, and the heterogeneous effects of patenting that vary by other characteristics of scientists or fields. Based on the empirical findings, Chapter three provides policy implications that help facilitate dissemination and utilization of scientific discoveries in government laboratories.

Chapter three focuses on demographic factors that influence mobility decision of inventors. First, it focuses on whether the distribution of the same surname in a given county influences the geographical mobility of inventors. Then, it examines how historic surname distribution and other factors, i.e., family-specific, inventor-specific, and institutional factors, interact to shape mobility decision of inventors. This provides consistent evidence that historic surname distribution predicts the geographical mobility of inventors and how this relationship may be altered by other factors.

Chapter four examines the impact of inventors on regional entrepreneurship. To resolve the endogeneity problem hindering the estimation of inventor effects, the chapter develops a shift-share instrument based on historic surname distribution in the US in 1940. This explains how the shift-share instrument may be used to create an exogenous variation for inventor mobility events. With the shift-share instrument, Chapter four estimates how inventor inflow influences the rate of startup foundation, quality of startups, and reallocation of startup activities across sectors. Moreover, it provides relevant policy implications and explains how the findings may help explain the agglomeration mechanism.

Chapter five concludes the dissertation by summarizing the chapters and suggests directions for future research.

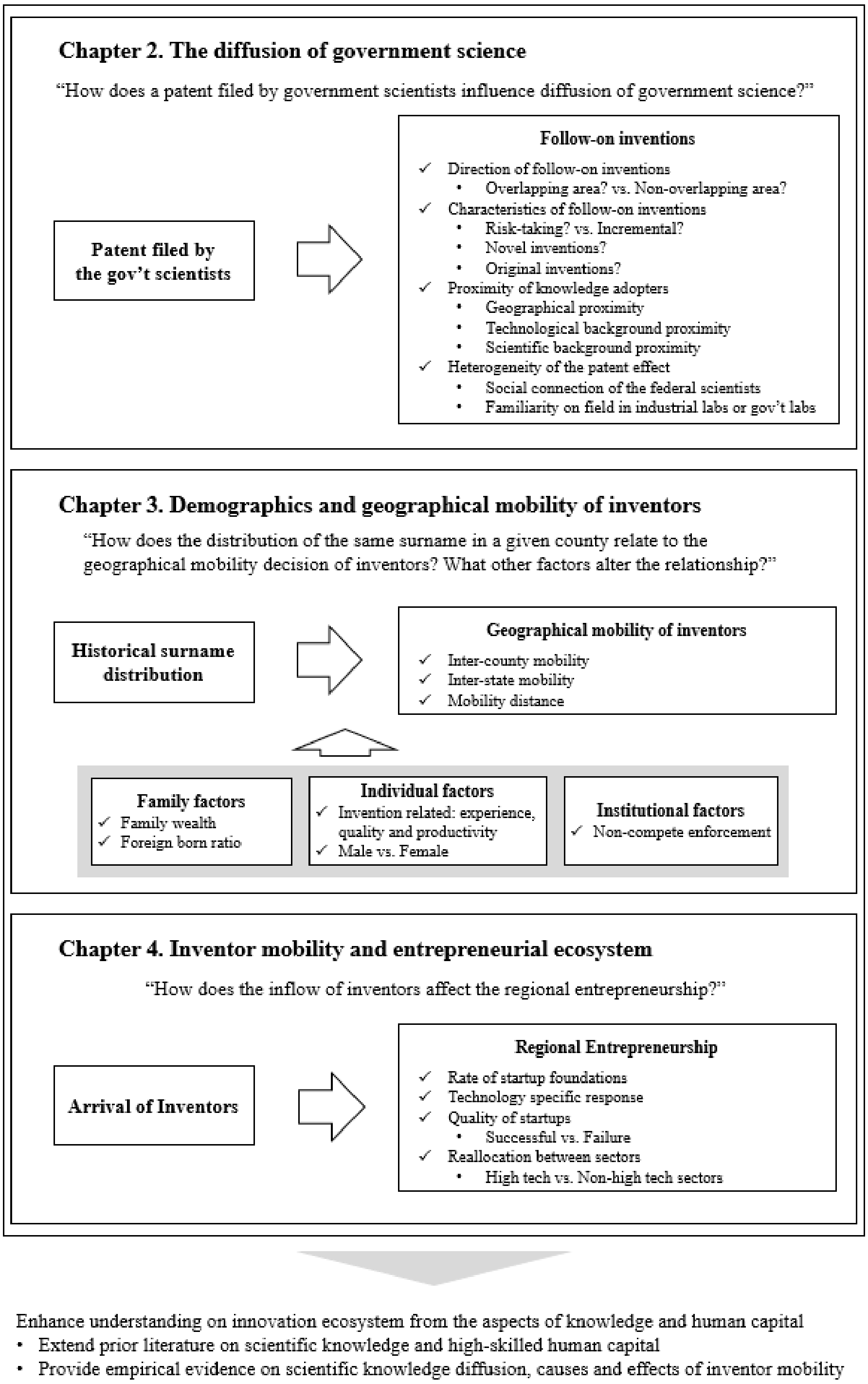


Figure 1-1. Research outline

Table 1-1, Research overview

	Chapter 2	Chapter 3	Chapter 4
Topics	The diffusion of government science	Demographics and mobility choice of inventors	Inventor mobility and startup formation
Focus	Scientific knowledge	Skilled human capital	Skilled human capital
Research Questions	How does a patent filed by government scientists influence diffusion of government science?	When do inventors move? Demographics and mobility decision	How does regional entrepreneurship respond to the inflow of inventors?
Methodology	Coarsened Exact Matching + Difference-in-Differences approach	OLS Regression + Interactions between factors	Instrumental variable approach (w/ Shift-share instrument)
Implication	Patents filed by government scientists help disseminate knowledge from government laboratories and increase follow-on inventions.	Historical surname distribution predicts geographical mobility of inventors.	Inflow of inventors increases not only the rate of startup foundation, but also the quality of startups.

Chapter 2. The diffusion of scientific discoveries in government laboratories[1]

2.1 Introduction

Scientific discoveries of government laboratories are a prominent source of industrial research and development (R&D). With substantial public investment—US federal laboratories, for instance, spent about $52 billion on research in 2017 (NSB, 2020)—government laboratories conduct basic science and applied research that is essential for technological advancement (Fleming & Sorenson, 2004; Jaffe & Lerner, 2001). The resulting scientific discoveries are transferred to the industry through direct pathways, such as cooperative research and development agreements (CRADAs), patent licenses and facility sharing (Adams et al., 2003). In addition to direct pathways, government laboratories also contribute to technological advancement through broader dissemination of scientific knowledge in the public domain (Bozeman et al., 2015; Mowery & Ziedonis, 2015). This broad knowledge dissemination of government science is important because firms in the industry often rely on external scientific knowledge to achieve further innovation (Arora et al., 2018; Poege et al., 2019). Importantly, government laboratories are also responsible for creating a broader societal impact with the full use of the results of public investment in research (Fini et al., 2018; Yin et al., 2021).

[1] This chapter is adapted from joint work with Junseok Hwang, Yura Jung, and Jisoo Lee. For the original article, please see the article titled "The diffusion of scientific discoveries in government laboratories: The role of patents filed by government scientists" (Conditionally accepted in Research Policy Journal as of January 1st, 2022)

Among various means that policymakers implement to ensure the full use of the scientific discoveries of government laboratories, we focus on the role of a patent filed by government scientists who participated in the research at the government laboratory. As patenting allows the protection of intellectual property rights as well as the commercial use of discoveries through licensing, government laboratories file patents on discoveries to appropriate their value, and policymakers have also encouraged such practices (Jaffe & Lerner, 2001; Lerner, 2002). Taking the United States as an example, the number of US patents filed has increased since the early 1980s as a series of policies were enacted to boost the transfer of knowledge from government laboratories through patents. In particular, the amendments enacted by the Federal Technology Transfer Act (FTTA) of the United States enacted in 1986 allowed scientists—including former employees who participated in the research at the time—to possess title to their discoveries at the government laboratory, resulting in a sharp increase in the number of patents filed by government laboratories (Jaffe & Lerner, 2001). Despite such encouragement of patenting activities, empirical evidence is lacking as to whether such activities facilitate or impede the adoption of government science in follow-on inventions.

Conceptually, the practice of scientists filing a patent on their discoveries in a government laboratory may have a double-sided effect on technological applications to follow-on inventions. To obtain this in more detail, we focus on the two distinctive aspects of government laboratories. First, scientific research becomes idiosyncratic or specialized within a government laboratory because of the long-term nature and persistency of research (Saavedra & Bozeman, 2004). Second, the accessibility of information on the technological applicability of government research for external researchers is often limited (Link et al., 2011). These characteristics act as barriers for other inventors, either in industrial laboratories or in universities, in their attempts to adopt science from government laboratories. Considering these characteristics, from a conceptual perspective, we argue that the double-sided aspects of patent protection, that is, the preemption of technological opportunity (Heller & Eisenberg, 1998; Galasso & Schankerman, 2015) and disclosure of technological information to the public (Czarnitzki & Toole, 2011; Graham & Hegde, 2015; Hegde & Luo, 2018), manifests depending on the relevance of the application areas and the characteristics of the inventors adopting the government science.

To estimate the effect of a patent filed by government scientists on follow-on inventions, we analyze a dataset of research papers published by scientists in US Federally Funded R&D Centers (FFRDCs) using a combination of the coarsened exact matching (CEM) technique with restrictive criteria and a difference-in-differences approach. We construct a matched sample comprising 740 research papers published by scientists from FFRDCs, i.e., 370 papers patented by government scientists and 370 counterfactual papers. When CEM matches, we match exactly on the laboratory (FFRDC), paper publication year, narrowly defined scientific field, and similar impact factor of the journal in which the paper is published. This allows us to compare government research that is patented to research not patented by government scientists but similar otherwise.

Then, adopting a difference-in-differences approach, we compare the rate of follow-on inventions based on scientific papers from government laboratories before and after a patent that was granted to government scientists relative to analogous changes observed in a counterfactual scientific paper.

We find that a patent filed by government scientists impedes follow-on inventions in technological areas overlapping with the focal patent but facilitates the rate of follow-on inventions in non-overlapping technological areas, consistent with our theoretical discussion on the double-sided effects of patent protection. Investigation of the increase in follow-on inventions in non-overlapping areas reveals that the follow-on inventions are mainly attributed to risk-taking inventions, that is, inventions with a high likelihood of resulting in either impactful or failure patents, rather than incremental inventions. Moreover, it shows that a patent filed by government scientists increases inventions with a high level of originality in particular. When we further examine the characteristics of inventors who adopt government science, we find that a patent filed by government scientists mainly increases inventions by inventors in distant locations in terms of both geographical and technological proximity. Based on the heterogeneity analysis by characteristics of government scientists and scientific fields, we find that the patent effect is pronounced when the government scientists involved in the focal discovery have fewer social connections and when the scientific field is less familiar in the industry. This provides empirical evidence on the one-sided mechanism that a patent filed by government scientists provide technological information or the potential of government science to the public and, thus, increases follow-on technological applications of government science.

This study provides valuable contributions to the academic literature on government laboratories, patent protection, and the relationship between science and technology. We extend the literature on government laboratories by focusing on the diffusion of government science. We shed light on the important role of a patent filed by government scientists in knowledge dissemination and add detailed nuance on how government laboratories may create a broader societal impact with their scientific discoveries. We also contribute to the literature on patent protection and the relationship between science and technology by adding empirical evidence on how government science links to technological applications via patents.

We provide policy implications conducive to the full use of scientific discoveries in government laboratories. We suggest that policymakers consider the potential overlap or relatedness with a certain area that they strategically aim to improve using government science when designing policies on the patenting activities of government scientists. We also suggest that policymakers not only leverage patents but also implement other means to increase the interaction between government scientists and industrial laboratories. Moreover, our findings may be used to motivate government scientists, as these findings alleviate the uncertainty of the impact of such scientists' long-term research in government laboratories. Finally, we emphasize the importance of establishing a structure that facilitates knowledge transfer from government laboratories to industrial firms, thereby creating synergy between the two types of organizations in response to the industrial landscape changes in the current era.

2.2 Related literature

This study contributes to and connects three distinct streams of research. First, it advances the literature on government laboratories by uncovering one of the mechanisms through which government science is disseminated and used in technological applications. Despite the importance of government laboratories as a pillar of the National Innovation System (Mowery, 1992; Nelson, 1993; Fagerberg & Srholec, 2008), research on the role of government laboratories has not been as prolific as research on the other two pillars, i.e., universities and firms in industry. Bozeman and Crow have examined the role of government laboratories focusing on technology transfer from government laboratories and related public policy in a series of studies (Bozeman & Crow, 1991; Crow & Bozeman, 1998; Bozeman, 2000; Bozeman et al., 2015). Jaffe and Lerner (2001) studied how patent policy on government laboratories changes over time and have demonstrated the trends in government laboratories' patent activities. Adams et al. (2003) showed that collaborative work between firms and government laboratories through CRADAs positively influences the research of the firms that adopt science of the government laboratory. In a relatively recent study, Fini et al. (2018) considers the mission of government laboratories for a broader societal impact and suggests the importance of effective scientific knowledge transfer into practical applications. We extend the prior research by suggesting patent protection on government laboratories' scientific discoveries as a significant institutional factor that determines scientific knowledge dissemination from government laboratories to other inventors' technological applications.

Our study also relates to studies on patent protection. The effects of patent protection on knowledge dissemination or follow-on invention is still on a debate among scholars. On one hand, scholars emphasize the positive aspect of disclosure of technological information through a credible and standardized channel. Studies suggest that the disclosure of technological information through patents allow to publicize inventions' existence and scope (Graham & Hedge, 2015), thereby increasing efficiencies in market for ideas (Gans et al., 2008; Hegde & Luo, 2018) and reducing duplicative follow-on research (Luck et al., 2020). On the other hand, scholars also suggest the negative aspect of the fragmentation of intellectual property rights, overlapping claims, and confinement of knowledge within the originating organization (Heller & Eisenberg, 1998; Ziedonis, 2004). Heller & Eisenberg (1998) suggest this as "the tragedy of the anticommons", in which describes patents hinder follow-on research and applications as many owners hold rights on prior discoveries. Galasso & Schankerman (2015) provides empirical evidence on this by showing that invalidation of a patent leads to an increase in follow-on inventions based on the focal patent. Melero et al. (2020) suggests that the legal protection of patent converts the individual participating inventors' expertise into the patent-holder-specific human capital, and adds empirical evidence that patents induce inventors to stay in the patent-holding organization. Extending such findings on the implication of patent protection in prior studies, we add empirical evidence on the debate by focusing on patent protection around scientific discoveries in government laboratories. This is particularly important as the mechanism by which government science is applied to follow-on inventions is distinctive relative to that of scientific knowledge generated in private industry or universities.

Finally, this study contributes to the literature on the relationship between science and technology. Science has long been emphasized as a key driver of technological innovation. It facilitates technological innovation by enabling effective management of R&D processes, reducing fruitless trial-and-error methods, and allowing novel and useful combinations of knowledge to form (Fleming & Sorenson, 2004; Owen-Smith, 2001). Indeed, researchers have presented empirical evidence that science facilitates technological innovation and enhances the quality of innovation through analyzing the relationship between journal papers and patents. Cockburn et al., (1999) emphasize how adoption of science-based drug discovery leads to higher R&D productivity and uncover that firms' organizational factors alter the rate of adoption, especially focusing on the pharmaceutical industry. Sorenson and Fleming (2004) show how science-based patents are more likely to be diffused and utilized by future inventions. Ahmadpoor and Jones (2017) found that more than 60% of patents link to a scientific paper, and patents with direct linkage to a scientific paper are found to be more useful in each field. A recent study by Poege et al. (2019) further investigates the impact of the quality of science on patent value, and shows a significant relationship between the quality of papers and value of patents. Government laboratories are a key source of scientific discoveries, especially those that involve persistent investment and time. Thus, it is important to demonstrate how inventors adopt and apply science for their technological inventions, focusing on scientific research conducted in government laboratories.

2.3 Background: Technology transfer and patent policy for US government laboratories

Since 1980, the US Congress has enacted a series of laws to promote technology transfer and scientific knowledge dissemination from US government laboratories to non-government laboratories, such as universities and R&D laboratories in the private sector. The Stevenson-Wydler Act enacted in 1980 was the first enactment that aimed to disseminate knowledge and information that originate in government laboratories by encouraging government laboratories to participate in technology transfer activities. For instance, it required government laboratories to spend a particular portion of their budget on technology transfer activities and to establish an Office of Research and Technology Application in laboratories with more than two hundred staff members. The ownership of intellectual property that results from government-funded research was addressed in the Bayh-Dole Act in 1980, which allowed government laboratories to retain rights to their inventions. Especially, the significance of the Bayh-Dole Act is emphasized from the various aspects in prior studies, e.g., on technology licensing and patenting (Mowery et al., 2001; Thursby & Thursby, 2003), commercialization of science (Kenney & Patton, 2009), and scientist entrepreneurship (Aldridge & Audretsch, 2011). The Federal Technology Transfer Act (FTTA) of 1986, which enacted several amendments to the Stevenson-Wydler Act, enables government laboratories to enter or participate in CRADAs with industry as well as negotiate licensing for their patented inventions developed in their laboratories. This legislation from the 1980s constitutes policy changes that shifted in favor of permitting exclusive rights for government science aiming to promote follow-on technological applications or commercialization (Bozeman, 2000; Adams et al., 2003).

Additional US legislation followed in the 1990s and 2000s that facilitated the diffusion of the scientific knowledge of government laboratories by amending regulations on intellectual property. For example, the American Technology Preeminence Act of 1991 included intellectual property as a potential contribution of the government laboratory under CRADAs, allowing intellectual property to be exchanged between the parties. The National Technology Transfer and Advancement Act of 1995 ensures that US firms obtain sufficient intellectual rights on inventions resulting from joint research under CRADAs with government laboratories. Furthermore, it increased the potential limit of payment in royalties, providing government scientists with financial incentives to develop research outcomes with commercialization potential. The Technology Transfer Commercialization Act of 2000 permitted technology licensing of existing inventions prior to a CRADA and allowed the granting of an exclusive or partially exclusive license to other organizations with early notification to the agency (Federal Laboratory Consortium for Technology Transfer [FLC], 2019).

Revising patent policy was also an important attempt to increase transfer or further use of the intellectual property of government laboratories. Notably, a number of amendments to the Stevenson-Wydler Act enacted by the FTTA of 1986 resulted in a substantial increase in patents filed by federal scientists. The FTTA enables current or former employees of government laboratories, such as scientists, engineers, or technical personnel, to possess title to the scientific discoveries that they make at the laboratory as well as to protect their intellectual properties and rights on discoveries by patenting the findings. Moreover, the Act requires royalty sharing[2] and rewarding[3] on licensed invention by law (Federal Laboratory Consortium for Technology Transfer [FLC], 2019; Federal Technology Transfer Act, 1986). These provide scientists of government laboratories with both financial and career incentives to file a patent for their invention made at the government laboratory, especially inventions that are considered impactful. As the statistics provided by Jaffe and Lerner (2001), the number of government laboratory patents sharply increases after 1986 (almost doubled from 1981 to early 1990s) while the R&D spending per fiscal year remains nearly constant.[4]

---

[2] FTTA specifies that the distribution of the royalties should not affect the regular payment of the employee, and that a fixed minimum reward system must be guaranteed. A certain percentage of the royalties received by the agency that exceeds a threshold amount is required to be shared with each inventor involved. This also stipulates that the total portion of royalty payments distributed to contributors shall exceed 15 percent of the total agency royalties of the fiscal year.

[3] FTTA requires the implementation of a reward system in each government laboratory to reward employees whose invention is highly recognized its value by invention activities of other institutes such as universities and firms in the private sector.

[4] Jaffe and Lerner (2001) provides statistics based on federally funded research and development centers (FFRDCs) operating under the US Department of Energy (DOE).

Table 2-1. US legislations to promote technology transfer

Year	Title	Description
1980	Stevenson-Wydler Technology Innovation Act	- Allowed dissemination of knowledge and information from federal laboratories - Required federal laboratories to take an active role in cooperation by allocating a portion of the budget for technology transfer activities - Required each laboratory to establish an Office of Research and Technology Applications (ORTA) to coordinate, and thus, facilitate technology transfer
1980	Bayh-Dole Act	- Addressed the issue of ownership of intellectual property rights, in terms of boundaries and licenses, arising from government-funded research - Allowed government owned and government operated (GOGO) laboratories to grant exclusive patent licenses to commercial organizations
1986	Federal Technology Transfer Act	- Made amendments to the Stevenson-Wydler Act - Enabled government laboratories to enter or participate in cooperative research and development agreements (CRADAs) - Allowed current or former employees to retain titles or receive royalty payments on inventions or licensed patents - Required implementation of reward systems in each government laboratory
1989	National Competitiveness Technology Transfer Act	- Enabled government-owned contractor-operated (GOCO) facilities to transfer technology into the private industries - Offered protection from disclosure to third parties
1991	American Technology Preeminence Act	- Included intellectual property as a potential contribution of the laboratory under CRADAs - Allowed exchange of intellectual property among the parties
1995	National Technology Transfer and	- Made amendments to the Stevenson-Wydler Act - Allowed US companies to have a sufficient intellectual property rights

	Advancement Act	on inventions resulting from a CRADA with a federal laboratory - Allowed US companies to choose an exclusive or nonexclusive license for inventions resulting from a CRADA with a federal laboratory
2000	Technology Transfer Commercialization Act	- Permitted federal laboratories to grant a license for a federally owned invention that was created prior to the signing of a CRADA - Required licensees to provide a plan for the application of the invention within a reasonable period of time
2007	America COMPETES Act	- Authorized programs in multiple agencies to fund basic research that involves high-risk - Eliminated government departments that were responsible for reporting and analysis of technology transfer activities

Note. Reorganized from Federal Technology Transfer Legislation and Policy ("The Green Book"), 6th Edition

2.4 Conceptual framework

We begin with a conceptual discussion on how scientific research in government laboratories has different characteristics, compared to scientific research from other types of research organizations. Given that the operation of government laboratories differs according to its purpose and that such laboratories conduct various types of scientific research, our conceptual discussion is not intended to be exhaustive. Rather, we pay particular attention to the two salient aspects of scientific research in government laboratories: 1) idiosyncratic knowledge accumulated within the government laboratory and 2) inadequacy of information on technological potential available to outside researchers. Building on prior research streams on government laboratories, patent protection and the application of science to technology, we explain the role of filing a patent around scientific discoveries by government scientists in follow-on technological applications of the scientific research.

2.4.1 Scientific research in government laboratories

Scientific research in government laboratories differs from that of other research organizations such as firms in industry or university laboratories in several respects. First, government laboratories conduct research that requires long-term and persistent investments and that aligns with the benefit of the public rather than a particular group (U.S. Department of Energy, 2020). This differentiates government laboratories from firms in the private sector that pursue scientific research that ultimately contributes to their technological advantage or product development and that alter their R&D direction in response to timely market needs (Arora et al., 2021). It also differentiates the role of government laboratories from universities that tend to conduct relatively more market-driven research activities (Link et al., 2011). Second, compared to firms that usually pursue scientific research by linking to their technological inventions and products or laboratories in universities, where the pressure of commercialization is more prevalent (Siegel et al., 2007), government laboratories are less prone to seek financial returns from their research. Although government laboratories also commercialize their scientific discoveries or technological inventions through partnerships (Jaffe & Lerner, 2001; National Academies of Sciences & Medicine, 2021) and the enactment of series of technology transfer-related bills, such as the Federal Technology Transfer Act (1986) and the National Competitiveness Technology Transfer Act (1989) in the United States, have encouraged commercialization, their research is not fundamentally driven by financial incentives per se, nor is it closely linked to commercial outcomes in order to compete in the market.

Such disparity in objectives and directions of research in government laboratories shape the distinct characteristics of their scientific discoveries. Due to the long-term nature and persistency of scientific research in government laboratories, relevant scientific knowledge and human capital have accumulated in government laboratories. This, in turn, leads to an increase in the idiosyncrasy and specialization of scientific knowledge within the originating government laboratories, which heightens the barriers for other scientists and inventors outside the laboratory when they attempt to utilize the scientific knowledge of government laboratories. Further, because scientific research in government laboratories centers on basic or applied research that was not originally intended for commercial use in the market and is less likely to be directed toward market needs (Link et al., 2011), information asymmetry exists in the application of scientific discoveries to follow-on technological inventions. That is, when scientists or inventors, either in industry or in academia, search for proper scientific knowledge in their invention process, they may find it difficult to recognize the scientific knowledge of government laboratories and its technological potential and applicability to their invention process.

2.4.2 Effect of patenting by government scientists on follow-on inventions

As reviewed in Section 2.2, prior literature suggests both positive and negative effects of patent protection in knowledge transfer or follow-on inventions, which we expect to remain when it comes to government science and patenting related to the discovery. When government scientists file patents on their discoveries, they are able to preempt promising technological opportunities related to the scientific discoveries. The preemption of opportunities by government scientists with specialized knowledge and complementary assets in the scientific research would reduce follow-on technological applications by other scientists or inventors, or, at a minimum, require radical, rather than incremental, application approaches. In regard to an overlapping technology area in particular, the additional transaction costs incurred by increased legal protection of government science may further discourage others from adopting and applying for their inventions (Heller & Eisenberg, 1998; Woo et al., 2015).

On the other hand, a patent filed by government scientists may help alleviate the inadequacy of information available to outside researchers. The patent filed by government scientists may inform the existence of scientific discovery and its technological potential for outside scientists or inventors (Graham & Hegde, 2015; Melero et al., 2020). In addition, it may contain information or guidelines regarding the technological application of the scientific discovery that government scientists may have perceived while participating in the process of the scientific research, thereby helping other scientists or inventors who are less familiar with the scientific research understand or perceive the technological potential of the research (Baruffaldi & Simeth, 2020). As much of the prior literature provides empirical evidence on the benefit of a patent on knowledge diffusion, follow-on technological applications, and the reduction of duplicative research (e.g., Baruffaldi & Simeth, 2020; Hedge et al., 2018; Luck et al., 2020), a patent filed by the government scientist may facilitate dissemination of the technological information or its potential to outside scientists or inventors who otherwise may find it challenging to adopt and apply the discovery in their inventions for a variety of reasons, such as distant geographical location, less relevant technological intellectual backgrounds, or the absence of social connections to government scientists.

Considering the double-sided aspects of a patent filed by government scientists, we expect that filing a patent would have different effects on follow-on inventions depending on the relevance of the areas in which the government science is applied as well as the characteristics of the inventors adopting the government science. In the following section, we empirically analyze how patenting by government scientists affects the rate of follow-on inventions using data on scientific discoveries in US federal laboratories.

2.5 Data and Empirical Specification

2.5.1 Data and Sample

Our primary data sources are the Microsoft Academic Graph (MAG) and Reliance on Science (RoS) databases. The MAG database provides bibliographic data on scholarly works, including authors, affiliations, and keywords (Sinha et al., 2015). The RoS database complements the MAG database with a novel dataset of patent citations to papers published since 1800. It also provides a crosswalk from the keywords of the MAG dataset to the scientific field classifications of other popular bibliographic databases such as the Web of Science and the OECD based on the probabilistic classification of the keywords (Marx & Fuegi, 2020). We gathered further data on each patent that cited at least one paper, including patent assignees and classification according to the Cooperative Patent Classification (CPC) system from the USPTO database. We complemented the patent data with the inventor and location data that was disambiguated and provided by Balsmeier et al. (2018).

We use data on research papers published by federal scientists affiliated with a US Federally Funded R&D Center (FFRDC). With a substantial amount of expenditure—accounting for 16.3% and 10.5% of the total R&D expenditures of the US government in 2018 and 2019, respectively, FFRDCs carry out advanced scientific research in various fields that is difficult for a single private organization or university to conduct, such as research in defense, health, security, and energy. They have also emphasized the transfer of their scientific and technological knowledge since the enactment of the Stevenson-Wydler Act in 1980, providing a suitable setting in which to analyze the role of patents in their scientific knowledge application to follow-on inventions. The empirical findings from the analysis will also provide relevant implications for US federal laboratories, specifically in terms of their patent policy.

To identify research papers published by US FFRDCs, we developed a rule-based text-matching algorithm to match the names of US FFRDCs from the raw affiliation strings provided by the MAG database. For the list of names of US FFRDCs, we used the master list of US FFRDCs, which is maintained by the National Science Foundation of the United States[5]. Our algorithm searches for the full or alternative names of each US FFRDC from the raw affiliation strings and identifies research papers that included federal scientists at US FFRDCs (see Appendix A1 for the detail). Out of a total of 324,028 unique papers identified as a paper published by at least one author from a US FFRDC, 215,843 papers were published in our sampling period from 1986 to 2013. Using this comprehensive data, we could show the research activities and direction of US FFRDCs, as shown in Figure 2-1. Further, Figure 2-2 demonstrates yearly trends of scientific research of US FFRDCs being applied to follow-on US patents, showing an increase in the diffusion of FFRDC scientific knowledge and technological application of FFRDC scientific research over time.

[5] The full list is available at https://www.nsf.gov/statistics/ffrdclist/

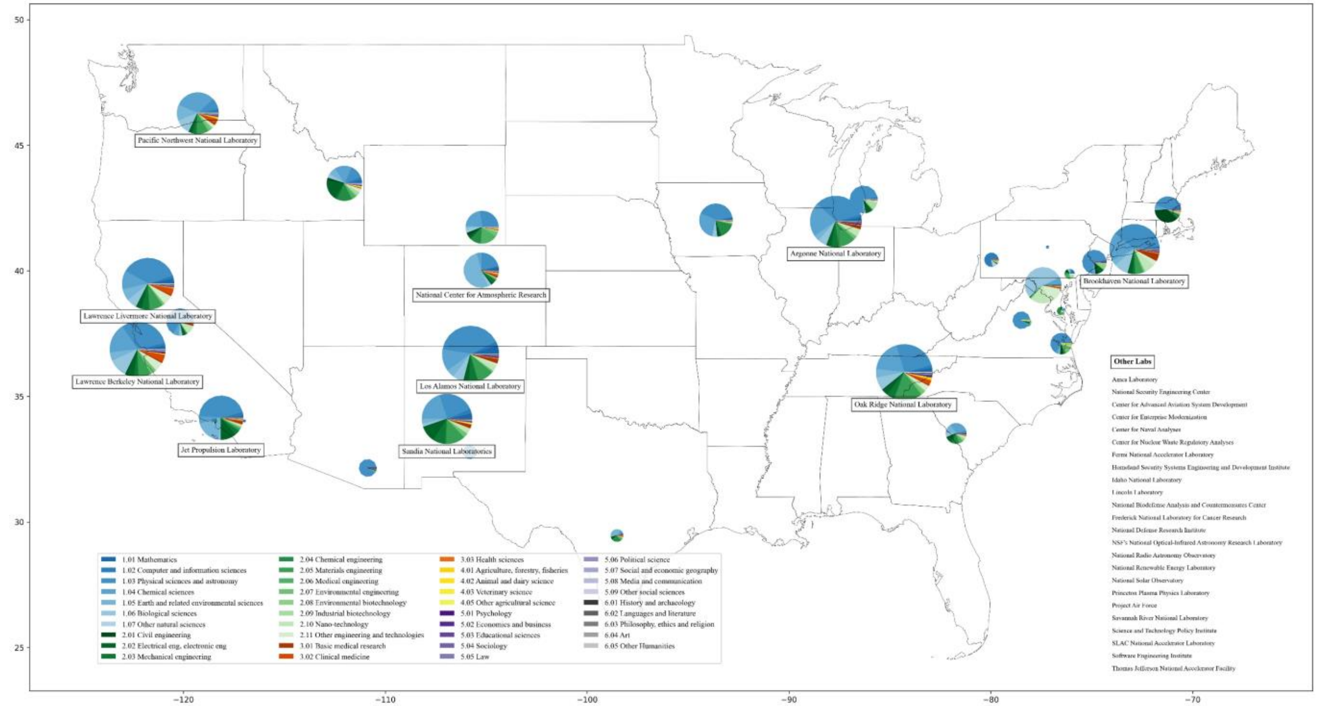


Figure 2-1. Research areas of US FFRDCs

Note: The choropleth shows the share of research fields of each US FFRDC. The scientific fields and paper shares are calculated based on papers published between 1986 and 2013 and identified as a paper of a scientist affiliated with an FFRDC by our rule-based text matching algorithm. The ten most prolific FFRDCs are labeled their name, whereas the remaining FFRDCs are listed on the side to avoid the complexity of the figure. Scientific papers are classified using the OECD subfields. The size of the pie chart of each FFRDC represents the amount of papers FFRDCs have published. Each pie chart is plotted based on the city location information of each FFRDC, while some locations are adjusted to avoid overlaps.

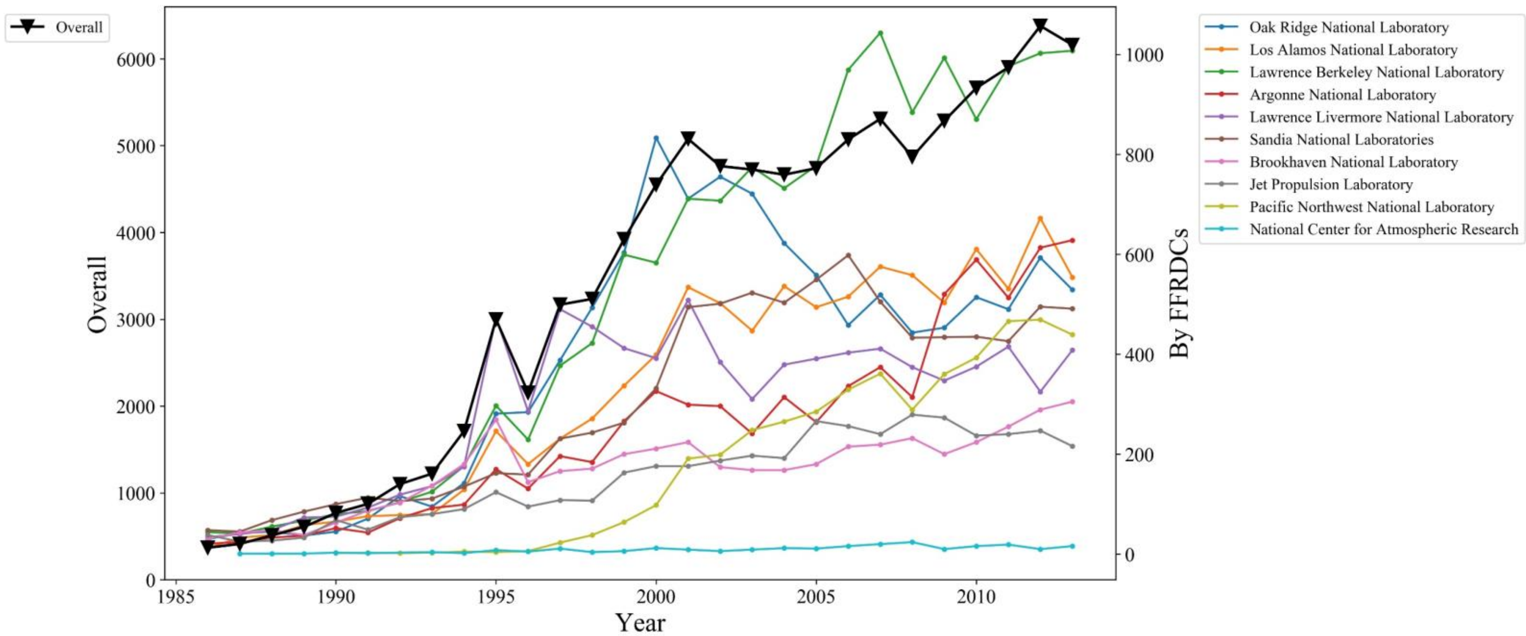


Figure 2-2. Increase in the number of US patents that build on scientific research of FFRDCs (1986 – 2013)

Note: We plotted the number of US patents citing at least one research paper of each FFRDC published within a ten-year window prior to the application year of each follow-on patent. We plotted for overall US FFRDCs as well as for each of the ten most prolific US FFRDCs throughout the sampling period.

In the estimations below, we used a subsample of the data on FFRDC research papers to rule out any confounding factors and measurement errors. As Adams et al. (2003) find, joint research with firms or external parties through CRADAs is a significant channel that influences research and knowledge transfer from federal laboratories. Therefore, we restrict papers to those that consist of only scientists affiliated with the US FFRDC organization, leaving us 79,733 unique papers. This allows us to alleviate the concerns that unobserved heterogeneity stemming from external collaboration partners confounds our estimation when comparing research papers. We also excluded papers with no citation received from a patent or a paper. This leaves us with 10,122 unique papers published by US FFRDCs.

Our identification strategy relies on a difference-in-differences approach using matched sample, which compares the rate of follow-on invention of research patented by the federal scientists who participated in the discovery versus research that is not patented by the federal scientists but otherwise similar. The approach of comparing samples matched based on similar underlying characteristics has the advantage of absorbing the effects of unobservable factors that otherwise may confound the estimation; thus, it has been adopted in many prior studies analyzing observational samples (see examples of similar patent comparison by Jaffe, Trajtenberg and Henderson (1993), Palangkaraya et al. (2011) and Moreira & Soares (2020) as well as a similar startup comparison by Polidoro Jr & Yang (2021)). We separate our treatment and counterfactual control groups of research papers published by US FFRDCs based on whether a paper has ever been patented by the scientists who were actually involved in the process of discovery in the federal laboratory. Following Marx & Hsu (2021), we tracked the names of inventors for all patents that cite each research paper and indicate papers that are cited in a patent containing an inventor of the same name. We then define the research papers that are patented by federal scientists as our treatment group and those that have never been patented by federal scientists as our counterfactual group. Out of the total 10,122 unique FFRDC papers, 2,674 papers (26.4%) in our sample were patented by the same federal scientist.

We then employed the coarsened exact matching (CEM) to match each paper with the most similar paper. The CEM technique improves the balance between papers in the treatment and control groups in terms of observable dimensions, reducing model dependence and causal estimation error (Iacus et al., 2009). Similar to the idea of using the twin design (Bikard & Marx, 2019), we employed exact matching for FFRDC organization, publication year, and keyword of the paper[6], accounting for the characteristics of organization, time, and specific scientific field. To account for the quality difference of each paper, we also include the yearly impact factor of the journal when each paper was published. Using the journal impact factor has the advantage of accounting for the quality of each paper, as this factor is not determined by the quality of one single paper, a factor that is typically criticized when using the number of forward citations as a measure of paper quality. We use discrete buckets, separated by quantile

[6] Microsoft categorized science papers using abstracts and titles of the papers. There are more than 200 thousand different subfields in this classification system as of 2020 MAG database.

between the minimum and maximum values (i.e., between 0 and 35.9). This procedure resulted in a matched sample comprising 740 papers (370 papers patented by federal scientists and 370 counterfactual papers), which originated from 14 US FFRDCs[7]. As we applied highly restrictive criteria to ensure that we included only papers that are matched with the most similar papers, that is, scientific research conducted by the same FFRDC in the same year in the same narrowly defined scientific field and published in journals of a similar quality, many paper observations and FFRDC organizations had to be dropped during the CEM matching process. Table 2-2 provides a summary of the observable characteristics of papers in each of the treatment and control groups before and after CEM matching.[8] The difference in mean values between the treatment and control groups decreased after CEM matching. The t-tests for post-matching also indicate that there is no statistically significant difference for any of the observable covariates between the two groups, except for the number of scientists involved. This alleviates the concern regarding endogeneity in comparing treatment and counterfactual control groups for government science.

[7] Scientific papers included in our matched samples are the scientific discoveries originated from the followin g US FFRDCs: Ames Laboratory, Argonne National Laboratory, Brookhaven National Laboratory, Idaho National Laboratory, Jet Propulsion Laboratory, Lawrence Berkeley National Laboratory, Lawrence Liver more National Laboratory, Lincoln Laboratory, Los Alamos National Laboratory, Frederick National Lab oratory for Cancer Research, National Renewable Energy Laboratory, Oak Ridge National Laboratory, P acific Northwest National Laboratory, Sandia National Laboratories

[8] Please see the notes in Table 1 for the definition of each variable.

Table 2-2. Overview of FFRDC papers' characteristics pre- and post-matching

	Patented by the federal scientist = 0			Patented by the federal scientist = 1			
	N	Mean	SD	N	Mean	SD	t-statistics
Panel A: Pre-CEM							
Total forward citations	7448	105.609	425.960	2674	109.719	351.785	-0.447
JIF	7448	2.051	2.157	2674	2.300	2.267	-5.041***
JCIF	7448	0.050	0.072	2674	0.056	0.071	-3.926***
Number of unique authors	7448	3.112	2.243	2674	3.946	2.582	-15.829***
Prior experience of authors	7448	23.659	34.355	2674	23.570	31.820	0.117
Self-citation ratio	7448	0.134	0.194	2674	0.152	0.197	-4.084***
Search breadth	7448	0.642	0.323	2674	0.658	0.310	-2.244**
Panel B: Post-CEM							
Total forward citations	370	94.816	167.915	370	133.811	464.787	-1.518
JIF	370	2.142	2.094	370	2.145	2.195	-0.022
JCIF	370	0.050	0.057	370	0.051	0.061	-0.290
Number of unique authors	370	3.311	2.392	370	3.943	2.725	-3.355***
Prior experience of authors	370	23.566	28.785	370	24.432	29.228	-0.406
Self-citation ratio	370	0.141	0.194	370	0.160	0.220	-1.281
Search breadth	370	0.615	0.326	370	0.630	0.335	-0.626

Note. Panel A and Panel B summarize observable characteristics for the papers before and after the CEM matching, respectively. In the last column, we report t-statistics and p-values of two-sample t-tests for equal means. ***, **, * denote significance levels of 1%, 5%, and 10%, respectively. Total forward citations indicate the total number of forward citations that the focal paper received by research papers; JIF and JCIF indicate the yearly journal impact factor and journal commercialization impact factor provided by RoS database, respectively; Number of unique authors indicates the unique federal scientists listed as an author of the paper; Prior experience of authors indicates the average of cumulative counts of previous papers published by the authors; Self-citation ratio indicates the ratio of self-citation to one of the authors' papers out of all backward citations; Search breadth indicates the extent of which the focal paper builds upon diverse fields of knowledge, measured by the proportion of unique narrowly defined scientific fields (i.e., MAG fields) in all backward citation counts.

To contrast the effect of patenting by the federal scientist, we created a yearly panel for five years before and after the grant year. We take five years for the pre-period because papers in our sample are first patented by the federal scientists within 5.93 years after the publication on average (median: 5 years).[9] We use the grant year of the focal patent to separate the pre- and post-periods as patents become effective beginning when they are granted and, thus, create the above-theorized effects on follow-on inventions. Our final sample has 6,939 paper-year observations. Note that some papers have a time gap between publication year and the grant year of a patent of the federal scientist that is shorter than five years. For these papers, we only include years since their publication year, resulting in an unbalanced panel.

Our main outcome variable of interest is the rate of inventions based on the focal FFRDC research paper. We count the number of patents applied for each year, which includes the focal FFRDC research paper in the list of the prior art. These patent citations to the focal research paper represent the knowledge flow from the science paper as well as the application of scientific discovery to a technological invention building upon the original scientific discovery (Roach and Cohen, 2013). To further analyze detailed aspects of follow-on inventions, we extend the measurement of patent counts by separating patents into a variety of bins that comprise related follow-on patents depending on the criteria of interest. Then, the follow-on patents in each bin are counted by aggregating the

[9] Our results are robust to alternative time windows.

patents at the focal paper and year levels. For instance, to analyze the differential effects of patenting on follow-on inventions in overlapping and non-overlapping technological areas, we separated the forward-citing patents into patents in overlapping and those in non-overlapping areas relative to the area of the focal patent granted to the federal scientist. Adopting the CPC classification of USPTO and using subgroup level classification, we separate patents with at least one overlapping subgroup and patents with no overlapping subgroups exclusively and count and aggregate the number of follow-on patents at the paper/year level to measure the rate of inventions in the overlapping and non-overlapping areas, respectively.

Summary statistics of our focal variables included in the analysis are shown in Table 2-3. FFRDC research papers included in our sample become a basis for 0.5 follow-on patents each year on average. Specifically, 0.21 follow-on patents with an overlapping subgroup are applied for each year, whereas 0.29 follow-on patents with no overlapping subgroups are applied for each year. The panel of papers in our sample spans from 1986 to 2018. The papers in our study are cited by around 7 academic papers each year. The journal impact factor and journal commercialization impact factor of journals in which the papers were published are 2.84 and 0.05 on average, respectively. Papers resulting from a collaboration with at least one author affiliated with a university account for 47% of research papers published by FFRDCs each year, suggesting that FFRDCs in our final sample collaborate actively with university scientists.

Table 2-3. Summary statistics

Variables	Mean	S.D.	Min	Median	Max
1. Overall follow-on patents	0.5	1.76	0	0	56
2. Follow-on patents with overlapping subgroups	0.21	0.93	0	0	37
3. Follow-on patents with no overlapping subgroups	0.29	1.33	0	0	48
4. Patented	0.5	0.5	0	1	1
5. Post	0.5	0.5	0	1	1
6. Forward citation received by research papers	6.96	29.04	0	2	873
7. FFRDC's collaboration with universities	0.47	0.12	0.08	0.48	0.74
8. JIF	2.84	3.24	0	2	49.88
9. JCIF	0.05	0.06	0	0.03	0.93

2.5.2 Empirical specification

We used a difference-in-differences estimation to compare the rate of follow-on invention of scientific discoveries in federal laboratories for which one of the federal scientists is granted a patent and those that are not patented. Because the difference-in-differences estimation estimates the treatment effect based on the difference in outcomes between treatment and counterfactual groups, it alleviates potential endogeneity and selection bias problems in the estimation. We estimate the following specification:

$$\mathrm{Y}_{\mathrm{i,t}} = \beta_1 Patented_i \times Post_t + \beta_2 Patented_i + \beta_3 Post_t + \gamma Z_{it} + \delta_t \times \theta_i + \rho_i + \tau_i + \epsilon_{\mathrm{i,t}} \quad (1)$$

where $\mathrm{Y}_{\mathrm{i,t}}$ represents the number of patents applied in year t building on the FFRDC research paper i. $Patented_i$ and $Post_t$ are the main explanatory variables of interest.

$Patented_i$ equals one for papers that have been patented by the federal scientist and zero otherwise. $Post_t$ equals one for the years after the patent is granted to the federal scientist. β_1 captures the effect of patenting by the federal scientist on follow-on inventions. Z_{it} is a set of time-variant control variables that may influence the outcome variables. To account for the quality or relevance of the paper in scientific research during the year, we control for the number of forward citations received from other scientific research published each year. We also include journal impact factors and journal commercial impact factors as control variables. Journals are important channels through which research papers are distributed; thus, journal effects should be controlled for. Finally, we include the extent to which FFRDCs collaborate with universities during the year as a control variable. This controls for potentially significant pathways through which the FFRDCs' knowledge spills over in either direct or indirect ways.

δ_t and θ_t denote a full set of year and scientific field fixed effects, respectively. We control for any time-invariant unobserved characteristics of scientific fields as well as for scientific field-specific trends by including scientific field-year fixed effects $\delta_t \times \theta_i$. ρ_i is a full set of US FFRDC fixed effects, which absorb the unobserved characteristics of the US FFRDC in which the scientific research was conducted. The inclusion of US FFRDC fixed effects mitigates the concern that the effect is driven by FFRDC organization-level characteristics. Finally, we include matched research papers fixed effect τ_i to effectively control for unobserved characteristics that the matched research papers share. This ensures that our identification comes mainly from variation within the matched research papers before and after a patent granted to the federal scientist.

We estimate the specification using Poisson regression as our main variables are non-negative and skewed variables. Specifically, we adopt the Poisson pseudo maximum likelihood (PPML) regression because it allows zero values to be handled without any adjustments, and its estimator is robust to overdispersion (Hausman et al., 1984). In addition, the standard errors in PPML are robust to serial correlation, alleviating the concerns raised in difference-in-differences estimation with panel data (Azoulay et al., 2019).

2.6 Results

2.6.1 Preliminary evidence of the rate of follow-on inventions after patenting by federal scientists

Descriptive statistics of our outcome variables in Table 2-4 show how the rate of follow-on inventions based on the federal research changes before and after a patent is granted to federal scientists. The change in the rate of follow-on inventions is moderate and not statistically significant if we count and aggregate all follow-on inventions based on the focal federal research (Panel A, DiD mean = -0.053, p = 0.527). However, this change becomes evident when we further separate the follow-on patents into patents with and without overlapping subgroups. The follow-on patents with subgroups that overlap with those of the patent filed by the federal scientist decrease after a patent is granted to the federal scientist relative to analogous changes observed among counterfactual research papers (Panel B, DiD mean = -0.187, $p < 0.01$). In contrast, follow-on patents with no overlapping subgroups increase after a patent is granted to the federal scientists compared to the counterfactual research papers (Panel C, DiD mean = 0.133, $p < 0.05$). This provides preliminary evidence for our core results on the effect of patenting related to discoveries at federal laboratories. We now turn to a regression framework to estimate the effect of patenting by federal scientists.

Table 2-4. Rates of follow-on patents for the patented and comparison discovery

	Panel A: Follow-on patents			**Panel B: Follow-on patents with overlapping subgroups**			**Panel C: Follow-on patents with no overlapping subgroups**		
	Before	After	*Diff*	Before	After	*Diff*	Before	After	*Diff*
Control	0.365	0.367	0.002	0.027	0.033	0.006	0.337	0.334	-0.003
Treatment	0.650	0.599	-0.051	0.478	0.297	-0.181	0.172	0.302	0.130
Diff	0.285	0.232	**-0.053**	0.451	0.264	**-0.187**	-0.165	-0.031	**0.133**

2.6.2 Main effect of patenting by federal scientists

Table 2-5 and Figure 2-3 show the estimation results on the effect of patenting by federal scientists on the follow-on inventions based on a scientific paper of an FFRDC, compared to the counterfactual similar scientific papers that are not patented by the federal scientists of the focal discovery. In Table 2-5, models (a) - (i) estimate Equation (1) for three different dependent variables that capture the rate of follow-on inventions, the count of all follow-on patents (models (a) - (c)), the count of follow-on patents with overlapping subgroups (models (d) - (f)), and the count of follow-on patents with no overlapping subgroups (models (g) - (i)). Models (a), (d), and (g) include only time-variant control variables that account for differences at the paper, journal, and FFRDC levels. Models (b), (e), and (h) include only the main variables, Patented, Post, and their interaction term Patented $\times$ Post, while models (c), (f) and (i) are the full models containing the main variables, the interaction term, and the control variables.

In model (c), no significant change in the rate of follow-on invention is found after a patent is granted to a federal scientist who participated in the discovery. Rather, an intriguing pattern is found when we further separate follow-on inventions into patents that are assigned one or more subgroups that overlap with the focal patent granted to the federal scientist and patents with no overlapping subgroups. Model (f) shows a stark decrease in the rate of follow-on inventions in the overlapping areas, compared to the counterfactual scientific papers. This finding is consistent with the view that federal scientists with specialized knowledge of the research may preempt the technological opportunities when they file a patent relating to their scientific discovery, discouraging other inventors from adopting federal science in their inventions.

Model (i) shows that the relative rate of follow-on inventions with no overlapping subgroups significantly increases after a patent is granted to the federal scientist of the focal discovery. In our preferred specification, model (i), the increase is estimated to be 66%[10]. This finding of the increased rate of follow-on inventions in non-overlapping areas supports the view that a patent filed by federal scientists informs researchers outside government laboratories of the existence of the government science and disseminates technological information and the potential of the government science to other inventors.

Figure 2-3 shows the dynamics of the effects of patenting by federal scientists on the rate of follow-on inventions with overlapping subgroups and with no overlapping subgroups. We interact the treatment effect with a set of indicators for each year relative to the grant year of the focal patent filed by the federal scientist of the focal discovery, and estimate a full model including all interacted terms. We used the coefficient obtained from the estimation for each year relative to the grant year of the focal patent and graphed for the rate of follow-on invention with overlapping subgroups and with no overlapping subgroups, along with a 95 percent confidence interval for each estimate (Panel A and Panel B in Figure 2-3, respectively). The graphs in Figure 2-3 show that the effect sizes increase over time. Particularly for the patents with no overlapping subgroups, Panel B in Figure 2-3 shows clear evidence that the rate of follow-on patents significantly increases after a patent is granted to the federal scientist of the focal discovery, and the increase lasts for the next five years, the timeframe that we use for the post-period in the sample.[11]

---

[10] $(e^{0.507} - 1) \times 100 = 66\%$.

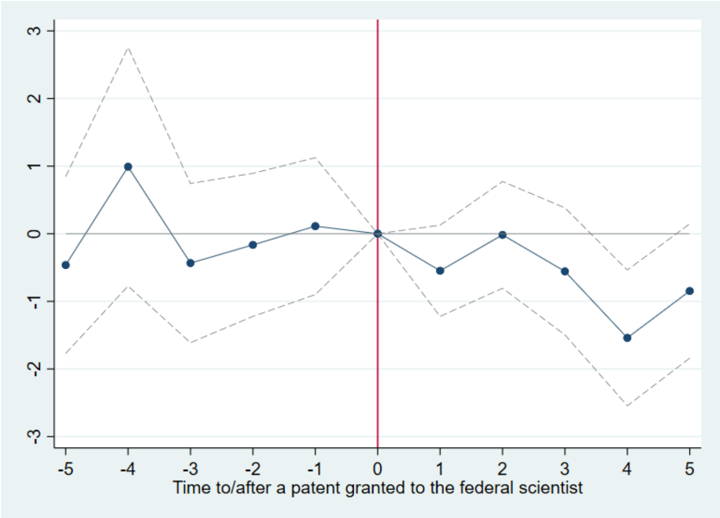
[11] The increase lasts up to the next seven years. See Appendix A1 for a graph with a longer post-period.

Table 2-5. Effect of patenting by the federal scientist on follow-on invention

	DV: Rate of follow-on inventions								
	Follow-on patents			Follow-on patents with overlapping subgroups			Follow-on patents with no overlapping subgroups		
VARIABLES	a	b	c	d	e	f	g	h	i
Independent variables									
Patented × Post		-0.029	-0.103		-0.714**	-0.710**		0.680***	0.507***
		(0.175)	(0.175)		(0.291)	(0.287)		(0.191)	(0.196)
Patented		0.581**	0.555**		2.913***	2.848***		-0.700**	-0.685***
		(0.234)	(0.218)		(0.415)	(0.406)		(0.278)	(0.243)
Post		-0.113	-0.030		0.181	0.235		-0.240	-0.151
		(0.153)	(0.150)		(0.276)	(0.274)		(0.227)	(0.222)
Control variables									
Forward citation received	0.008		0.008	0.006*		0.003	0.010**		0.010**
by research papers	(0.006)		(0.006)	(0.003)		(0.003)	(0.005)		(0.005)
FFRDC's collaboration	-1.884*		-1.890*	-1.774		-2.005	-2.022		-2.127
with universities	(1.058)		(1.045)	(1.378)		(1.406)	(1.510)		(1.548)
JIF	0.005		0.010	-0.021		0.023	0.037		0.035
	(0.033)		(0.033)	(0.028)		(0.040)	(0.037)		(0.037)
JCIF	2.854***		2.414***	4.051***		2.345***	1.629		2.001**
	(0.686)		(0.654)	(0.783)		(0.876)	(1.023)		(1.019)
Observations	6,939	6,939	6,939	6,939	6,939	6,939	6,939	6,939	6,939
Number of Matched science pair	331	331	331	331	331	331	331	331	331
Science Field - Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

Panel A. Follow-on patents with overlapping subgroups



Panel B. Follow-on patents with no overlapping subgroups

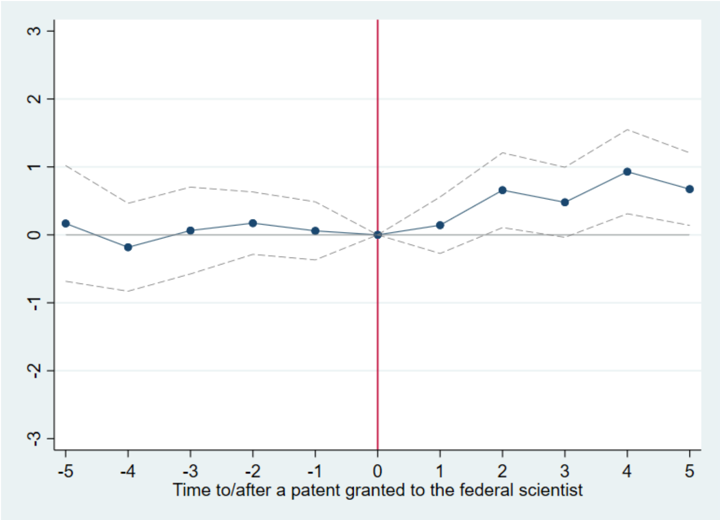


Figure 2-3. Effect of patenting by federal scientist on follow-on inventions

Further, as an ex-post validation of our identification strategy, the graphs in Figure 2-3 suggest that DiD estimators are not significantly different from zero before a patent is granted to the federal scientist of the focal discovery.

2.6.3 Characteristics of follow-on inventions based on federal science

In the remainder of the manuscript, we attempt to uncover the characteristics of the follow-on inventions based on federal science after a patent is granted to the federal scientist of the focal discovery. As we find in the previous section, follow-on inventions mainly decrease in the overlapping area and increase in non-overlapping areas. Thus, we focus on follow-on patents with no overlapping subgroups and exclude all patents that are assigned one or more overlapping subgroups with the focal patent filed by the federal scientist when counting patents for the dependent variables below.

Impact of follow-on inventions. —We first seek to uncover how patenting by the federal scientist affects the quality aspect of the follow-on inventions and infer the research direction of the follow-on inventions from the results. To do so, we follow recent studies that consider the distribution of forward citations received in analyzing the impact of patent or publication outcomes and provide nuanced details on whether follow-on patents are incremental or risk-taking inventions (e.g., Balsmeier et al., 2017; Azoulay et al., 2019). We split the follow-on patents into five exclusive bins based on the distribution of total forward citation counts among all science-based US patents and count the patents to measure the five outcome variables[12]: (1) the number of patents that fall into the bottom quintile of the forward citation distribution as well as failed patents with no forward citations; (2) the number of patents that fall into the second quintile; (3) the number of patents that fall into the third quintile; (4) the number of patents that fall into the fourth quintile; and (5) the number of patents that constitute the highest quintile. Models (a) – (e) in Table 2-6 report the estimation results for each outcome variable. We find that patenting by federal scientists significantly increases the number of follow-on patents that fall into the highest quintile as well as patents that failed or fall into the bottom quintile, while no evidence is found with regard to the number of follow-on patents that fall into the 2nd to 4th quintiles. We further separate the highest quintile into (6) patents that fall between the 80th and 90th percentile and (7) patents above the 90th percentile, for which models (f) and (g) in Table 5 report the estimation results. We find a significant change in the number of follow-on patents that fall between the 80th and 90th percentiles. If we consider patents that fall into the highest and lowest quintiles as risk-taking inventions following Balsmeier et al. (2017), we can interpret these results as indicating that patents filed by federal scientists mainly influence risk-taking type of inventions rather than the incremental type of inventions.

[12] We present the results when we use the distribution within each CPC group and application year. We also find consistent results when we use distribution at a coarsened level, e.g., distribution within each CPC subsection and application year.

Table 2-6. Impact of follow-on inventions

	Long run citation quintile						
	Failed and Bottom quintile	2nd quintile	3rd quintile	4th quintile	5th quintile	Btw. 80th to 90th percentile	Above 90th percentile
	a	b	c	d	e	f	g
Patented × Post	0.689**	0.647	0.608	0.286	0.557**	0.740**	0.408
	(0.308)	(0.504)	(0.395)	(0.318)	(0.245)	(0.337)	(0.322)
Observations	6,939	6,939	6,939	6,939	6,939	6,939	6,939
Number of Matched science pair	331	331	331	331	331	331	331
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Science Field-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. All regressions include independent terms of *Patented* and *Post* as well as all other controls. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

Novelty and originality of follow-on inventions. —We turn to see how patenting by the federal scientist of the focal discovery is linked to the novelty or originality of follow-on inventions. Along with the impact of an invention, novelty and originality are key dimensions that determine an invention's quality (Arts & Fleming, 2018; Jung and Lee, 2016). To capture the novelty of each follow-on patent, we used pairwise subgroup recombination following the existing literature (Arts & Fleming, 2018; Jung & Lee, 2016; Uzzi et al., 2013). We separated follow-on patents into two exclusive bins: (1) patents containing at least one pairwise subgroup recombination that had not previously existed among science-based patents in the USPTO database and (2) patents without such pairwise subgroup recombination. Models (a) and (b) in Panel A of Table 2-7 report the estimation results for the outcome variables. They suggest that patenting by federal scientists increases follow-on inventions both with and without novel recombination, but their effect sizes are not significantly different. Models (c) and (d) in Panel A of Table 2-7 report the results from the same estimation but consider the first appearance of pairwise subgroup recombination among all patents in the USPTO database rather than simply the science-based patents. We find similar results here, suggesting that the effect of patenting by federal scientists does not differ based on the novelty of follow-on inventions.

To measure the level of originality of each follow-on patent, we adopt the originality measurement based on the Herfindahl-Hirschman Index (HHI) following Hall, Jaffe, and Trajtenberg (2001). Referring to the backward citations that each patent made to the prior art, we calculate the originality of each patent using the following equation: $1 - \sum_{i}^{n_i} {s_i}^2$, where s_i represents the share of each CPC group i out of all CPC groups assigned to the cited patents, and n_i represents the number of unique CPC groups assigned to the cited patents. We then split the follow-on patents based on the median value of originality based on the distribution of all science-based US patents. Models (a) and (b) in Panel B of Table 2-7 report the results for follow-on patents with above median and below median originality, respectively. We find a positive and significant effect of patenting by federal scientists on follow-on patents with above-median originality, whereas no evidence is found for any changes in the number of follow-on patents with below-median originality. We also find consistent results in models (c) and (d), where we split the follow-on patents using the median value of originality based on the distribution of patents applied in the same year only. This can be interpreted as indicating that the increase in follow-on inventions following patents granted to federal scientists is mainly attributable to patents with a high level of originality.

Table 2-7. Novelty and originality of follow-on inventions

	New-to-the-world subgroup recombination (among science-based patents)		New-to-the-world subgroup recombination	
	Patents with novel recombination	Patents without novel recombination	Patents with novel recombination	Patents without novel recombination
	a	b	c	d
Panel A: Novelty				
Patented × Post	0.498**	0.593***	0.468*	0.623***
	(0.253)	(0.214)	(0.258)	(0.219)
	Originality based in HHI (Median split using distribution of overall patents)		Originality based in HHI (Median split using distribution of patents applied in the year)	
	Patents with above-median originality	Patents with below-median originality	Patents with above-median originality	Patents with below-median originality
	a	b	c	d
Panel B: Originality				
Patented × Post	0.443**	0.171	0.427**	0.221
	(0.204)	(0.280)	(0.203)	(0.278)
Observations	6,939	6,939	6,939	6,939
Number of Matched science pair	331	331	331	331
Other controls	Yes	Yes	Yes	Yes
Science Field-Year FE	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes

Note. All regressions include independent terms of *Patented* and *Post* as well as all other controls. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

2.6.4 Proximity of knowledge adopters

In the previous sections, we demonstrate how patenting by federal scientists changes the rate of follow-on inventions as well as how it affects the detailed characteristics of follow-on inventions. Considering the aspect of knowledge adopters, that is, scientists or inventors who apply the focal federal science in their inventions, we examine how the effect of patenting by federal scientists differs depending on the proximity of the knowledge adopters to the focal federal science. Here, we examine three aspects of proximity: geographical proximity, technological proximity, and scientific intellectual proximity. Because patents may have multiple inventors, we consider the information of all inventors involved in each patent when we capture the proximity of knowledge adopters to the focal federal science.

To examine geographical proximity, we use the boundaries of states in the United States and separate between (1) follow-on patents applied for by inventors in the same state as the state of the FFRDC from which the focal discovery originated and (2) follow-on patents applied by inventors in different states. Models (a) and (b) in Table 2-8 report the estimation results, showing that patenting by federal scientists significantly increases the rate of follow-on inventions by inventors located in different states, while it does not have a significant effect on inventors in the same state. This finding is consistent with our conceptual discussion that patenting by federal scientists helps diffuse information on the focal discovery to inventors that otherwise may not be able to access such knowledge from federal laboratories; specifically, it supports the role of patenting in knowledge diffusion across geographical boundaries.

For technological proximity, we compared the set of subgroups in which inventors' previous patents were assigned with the focal patent granted to the federal scientists. We separate between (1) follow-on patents applied for by inventors with overlapping subgroups in their previous patents and (2) follow-on patents applied for by inventors with no overlapping subgroups in their previous patents. Models (c) and (d) in Table 2-8 report the results. Similar to the results on geographical proximity, these results show that patenting by federal scientists significantly increases the rate of follow-on inventions by inventors with no overlapping subgroups but not inventions by inventors with overlapping subgroups. This finding supports the role of patenting by federal scientists in knowledge diffusion, particularly, across technological boundaries.

We also examine the proximity in the scientific intellectual background of inventors, as compared to the focal discovery (models (e) to (f) in Table 2-8) or the FFRDC from which focal discovery originates (models (g) to (h) in Table 2-8). We capture inventors' scientific intellectual background using narrowly defined scientific fields (i.e., MAG field) of papers that their previous patents build upon, and compare them to the scientific field of the focal discovery or all scientific fields in which the FFRDC has publications. We find that the effect of patenting by federal scientists is significant for follow-on inventions regardless of the proximity in scientific intellectual background and that the effect sizes are not significantly different.

Table 2-8. Proximity of knowledge adopters

	Geographical proximity		Technological Intellectual proximity to the focal patent	
	Inventor team located in the same state	Inventor team located in different states	Inventor team with overlapping tech background	Inventor team with no overlapping tech background
	a	b	c	d
Patented × Post	0.048	0.583***	0.158	0.489*
	(0.270)	(0.212)	(0.298)	(0.250)
	Scientific Intellectual proximity to the focal discovery		Scientific Intellectual proximity to the focal FFRDC	
	Inventor team with overlapping scientific background	Inventor team with no overlapping scientific background	Inventor team with overlapping scientific background	Inventor team with no overlapping scientific background
	e	f	g	h
Patented × Post	0.487*	0.474**	0.417*	0.800**
	(0.256)	(0.221)	(0.219)	(0.318)
Observations	6,939	6,939	6,939	6,939
Number of Matched science pair	331	331	331	331
Other controls	Yes	Yes	Yes	Yes
Science Field-Year FE	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes

Note. All regressions include independent terms of *Patented* and *Post* as well as all other controls. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

2.6.5 Heterogeneous effect of patenting by federal scientists

We now turn to uncovering the heterogeneous effect of patenting by federal scientists along a number of dimensions with regard to federal scientists and scientific fields. First, we test whether the patenting effect differs by the extent to which federal scientists have established social networks with other scientists. To measure the social networks of federal scientists, we count the number of unique co-authors with whom federal scientists have collaborated up to ten years prior to the publication year of the focal discovery. Models (a) and (b) in Table 2-9 report the results for estimation using samples below and above the median value of the social networks of federal scientists. We find evidence for follow-on inventions increases for the samples below the median, whereas no evidence is found for the samples above the median. This is consistent with our expectation because the role of patents filed by federal scientists is amplified when detailed technological information is hardly available otherwise. If federal scientists are already well connected with other scientists, tacit knowledge around the discovery is more likely to be transferred via social networks, diluting the effect of patenting by federal scientists.

We then uncover how the effect of patenting by federal scientists differs according to the characteristics of each field. Specifically, we examine the extent to which research from industrial laboratories (models (c) and (d) in Table 2-9) or government laboratories (models (e) and (f) in Table 2-9) accounts for the research activities in the field. These estimations are motivated by the idea that scientific fields with which industrial laboratories are not familiar require further guidance to adopt and apply the discoveries from federal laboratories to their technological inventions.

To capture the contribution of industrial and government laboratories in each scientific field, we used the MAG field classification and separately calculated the share of scientific papers that originated from industrial and government laboratories for each field. First, the origin of science was extrapolated based on author affiliation information provided by the MAG database. We developed a rule-based text-matching algorithm to classify affiliation types using raw strings provided by the MAG database (see Appendix A1 for details). Papers were classified as science originating in an industrial (government) laboratory if our algorithm identified one of the authors as affiliated with an industrial (government) laboratory. Then, we grouped these science papers by the MAG science field classification and each decade[13], then we calculated the proportion of papers originating in industrial (government) laboratories. The overall average share of papers from industrial laboratories is 6.3%, and that of paper from government laboratories is 17.7% for the fields included in our sample.

Models (c) and (d) in Table 8 report estimations with samples split based on the median value of the share of research from industrial laboratories. Models (e) and (f) in Table 2-9 report estimations with samples split based on the median value of the share of research from government laboratories. The results show that the rate of follow-on inventions increases following patents filed by federal scientists if the scientific area is not familiar among industrial laboratories. Again, this supports our logic that patenting by

[13] Since the MAG field is a fine-grained classification based on the keywords, many fields scarce papers to calculate the contributions by industrial and government laboratories if we calculate the contributions for each year. This led us to aggregate the papers by each decade.

federal scientists benefits other inventors, especially in the industry, by providing technological information to those who are not familiar with the scientific field.

Table 2-9. Heterogeneous effect of patenting by the federal scientists

	Social network with other scientists		Share of scientific research from industrial research		Share of scientific research from government laboratories	
	Below median	Above median	Below median	Above median	Below median	Above median
	a	b	c	d	e	f
Patented × Post	0.511**	0.416	1.002***	-0.124	0.254	0.200
	(0.238)	(0.266)	(0.230)	(0.242)	(0.344)	(0.225)
Observations	3,452	3,487	3,456	3,483	3,466	3,473
Number of Matched science pair	232	236	255	264	268	267
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Science Field-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes	Yes	Yes

Note. All regressions include independent terms of *Patented* and *Post* as well as all other controls. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

2.6.6 Placebo tests

To further enhance the robustness of our results, we ran two placebo tests. Models (a) and (b) in Table 2-10 present the results from the estimation in which we include only papers for which the first patent built on the paper was not applied for by (model (a) in Table 2-10) or not granted to (model (b) in Table 2-10) one of the federal scientists who participated in the process of the respective scientific research. No evidence was found on the effect of patenting by federal scientists from these estimations. This alludes that providing guidance and information regarding the discovery in federal laboratories is among the key mechanisms by which patenting by federal scientists affects follow-on inventions, and such an effect is valid only when one of the federal scientists who actually participated in the scientific research at the federal laboratory files a patent related to the discovery. Models (c), (d), and (e) in Table 2-10 report the estimation of overall follow-on scientific research, only follow-on scientific research published in the same field, and only follow-on scientific research published in different fields, respectively. Similar to follow-on inventions, we used paper-to-paper citation data to capture follow-on scientific research that cites the focal paper in our sample. As discussed above, a patent filed by federal scientists either preempts technological opportunities or provides information on technological potential to subsequent inventions. However, this should matter only for technological applications and not for scientific research activities. Consistent with this logic, we find no evidence that patenting by federal scientists influences follow-on scientific research or its direction.

Table 2-10. Placebo tests

	Effect of patenting by other scientists		Follow-on scientific research		
	Exclude papers that the first patents applied for by the author	Exclude papers that the first patents granted to the author	Overall	Same field	Different field
	a	b	c	d	e
Patented × Post	0.046	0.231	0.090	0.158	0.071
	(0.227)	(0.276)	(0.066)	(0.145)	(0.064)
Observations	3,397	3,582	6,939	6,939	6,939
Number of Matched science pair	289	289	331	331	331
Other controls	Yes	Yes	Yes	Yes	Yes
Science Field-Year FE	Yes	Yes	Yes	Yes	Yes
FFRDC FE	Yes	Yes	Yes	Yes	Yes
Matched science pair FE	Yes	Yes	Yes	Yes	Yes

Note. All regressions include independent terms of *Patented* and *Post* as well as all other controls. Standard errors clustered at the matched science-pair level are reported in parentheses. ***, ** and * denote a significance level of 1%, 5%, and 10%, respectively.

2.7 Discussion and Conclusion

By uncovering the effect of a patent filed by government scientists on follow-on inventions, this study makes several contributions to three different areas of research. First, it contributes to the literature on government laboratories by advancing the understanding of the dissemination of scientific knowledge from those laboratories.

Extending prior studies focused on the full use of scientific knowledge or technology transfer from government laboratories via various channels (see Section 2 for a detailed literature review), the findings of this study suggest patent protection as a significant institutional means that influences the dissemination of government science. Specifically, the study shows how filing a patent on a scientific discovery in a government laboratory may assuage the distinctive characteristics of scientific research in government laboratories, that is, the idiosyncrasy of scientific knowledge and information inadequacy to outside researchers, which otherwise may act as a barrier for other inventors when attempting to adopt government science. This adds detailed nuance regarding how government laboratories can create broader societal impact with their scientific discoveries, which recent studies in the field suggest as an important mission of government laboratories (e.g., Bozeman et al., 2015; Fini et al., 2018)

Second, the study advances the literature on patent protection by adding empirical evidence on how filing a patent on a scientific discovery affects follow-on inventions. Our results suggest that the double-sided effects of patent protection, that is, preemption of technological opportunity vs. disclosure of technological information to the public (see Section 2 for a detailed literature review), also apply to the patents filed on government science. While follow-on inventions decrease in the overlapping areas due to the preemption of the key opportunity related to government science by the focal patent, they increase in the non-overlapping areas as the focal patent provides guidance on technological application or the potential of government science. We add empirical evidence of the patent effect to the literature, specifically focusing on scientific knowledge discovered in government laboratories.

Third, this study contributes to the literature on the relationship between science and technology. Extending prior studies that emphasize science as a key source of technological innovation (see Section 2 for a detailed literature review), this study uncovers the underlying mechanism by which science in government laboratories is adopted and applied to follow-on inventions. In particular, it provides empirical evidence on how a patent filed by government scientists influences risk-taking and original follow-on inventions. In addition to prior studies that suggest how science contributes to technological inventions (e.g., seminal study, Fleming and Sorenson (2004)), this finding suggests that the filing of patents related to a scientific discovery by the scientist who participated in the discovery may facilitate the full use of the focal scientific knowledge in its application to follow-on technological inventions.

Along with the contributions to the academic literature, the findings of the study suggest invaluable policy implications. Policymakers should be aware of the double-sided effects of a patent filed by government scientists when designing policies that regulate government laboratories and scientists. Our findings demonstrate that a patent filed by government scientists leads to a decrease in follow-on inventions in overlapping areas, while it increases follow-on inventions in non-overlapping areas. Moreover, a patent filed by government scientists increases follow-on inventions characterized by risk-taking and high originality, and it benefits inventors who are not (geographically or technologically) close to government science. Policymakers may exploit such findings and implement patent policies according to their strategic needs. For instance, policymakers often want to make improvements in particular technological areas for a national good, such as the United States Innovation and Competition Act of 2021, where US policymakers strategically aim to foster basic and applied technology research in certain strategic areas, such as artificial intelligence, semiconductors, and biotechnology. In such cases, scientific research conducted by government laboratories with relevant capabilities may become crucial ground for follow-on inventions in industrial research. Thus, policymakers can implement patent policy accordingly, that is, either encourage or discourage patenting on government science, considering the potential overlap in technological fields with follow-on inventions in areas in which policymakers are willing to improve, thereby reducing any friction caused by the patents of government scientists and leveraging the benefit of technological information provided by the patents.

The findings suggest that policymakers should not only leverage patents but also implement other means to increase the use of government science. A patent filed by government scientists increases the adoption of government science by inventors located in distant areas in terms of geographical and technological proximity. In addition, it has been found to increase the follow-on inventions when the government scientists of discovery have fewer social connections to other scientists as well as when industrial laboratories are less familiar with the focal scientific field. That is, technological information disclosed to the public via patents facilitates the follow-on use of government science by other inventors whose access to the focal government science would otherwise be limited. On the other hand, the results show no evidence of the patent effect for closely located inventors, socially well-connected scientists, and fields already familiar to industrial laboratories. This alludes to the importance of opportunities to establish social networks as well as to share internal knowledge of government laboratories with inventors in the industry, suggesting that policymakers should devise and implement other means to increase interaction between government scientists and industrial laboratories.

The findings of the study can be used to motivate government scientists to strive in their scientific research and to file patents related to their discoveries in government laboratories. Filing a patent itself benefits government scientists by allowing them to receive royalties on their inventions when their patented inventions are used in any subsequent inventions or product development. However, in addition to financial rewards, government scientists may also be motivated by how their research in government laboratories could be used and how it can have a broader impact on technological advancement or any practical uses. This is important because government scientists often face a sizable chasm between scientific research and technological applications. Combined with the nature of research projects in government laboratories that involve long-term investment, an unclear path forward may lead government scientists to become doubtful about their research conducted in government laboratories. The study provides important evidence that their scientific discovery indeed influences follow-on technological inventions, especially when it is patented, thereby encouraging research in government laboratories.

Finally, this study also provides timely and essential implications for the industrial landscape of the current era. Scientific or technological areas, such as space exploration, energy, and nuclear power, have been deemed as areas in which only government laboratories control all roles from basic research to the implementation of the relevant science for practical use. However, as many firms in the private sector diversify their business and startups jump into such fields, the role of commercialization or implementation for practical use is shifting toward the private sector. Taking the example of the space industry, Space Exploration Technologies Corp. (SpaceX) strives to develop spaceships operating with a recyclable rocket booster[14], and Blue Origin recently carried out a successful launch of their spaceship[15], opening the door for the possibility of space tourism. In the energy sector, incumbent firms or startups attempt to develop technologies to replace traditional sources of energy with clean energy. For instance, TAE Technologies Inc. and General Fusion Inc., which are backed by Google[16] and Jeff Bezos from Amazon.com, Inc.[17], respectively, strive to develop nuclear fusion technologies and turn them into viable energy sources.

While such a shift is occurring in the industrial landscape, government laboratories

---

[14] https://www.reuters.com/article/us-space-spacex-launch-idUSKBN1711JY
[15] https://www.reuters.com/technology/jeff-bezos-worlds-richest-man-set-inaugural-space-voyage-2021-07-20/
[16] https://www.reuters.com/technology/google-backed-nuclear-energy-firm-tae-technologies-raises-280-mln-2021-04-08/
[17] https://www.bbc.com/news/science-environment-57512229

and industrial firms may complement each other. A recent study by Arora et al. (2018) provides empirical evidence that industrial firms reduce scientific research and rely instead on scientific knowledge from external sources (Arora et al., 2018). Thus, scientific research in government laboratories equipped with relevant capabilities is imperative to boost advancement in such areas. This calls for the establishment of a structure that facilitates knowledge or technology transfer from government laboratories to industrial firms to create synergy between organizations. Although this study provides only evidence on the diffusion of scientific knowledge facilitated through patents filed by government scientists, it sheds light on how scientific knowledge of government laboratories is diffused and who may benefit from it. Further investigation on various determinants or conduits that facilitate the diffusion of government science is warranted in future research to establish such complementary structures.

Chapter 3. Demographics and geographical mobility of inventors[18]

3.1 Introduction

The geographical mobility of inventors has received attention in the literature on inventor mobility. As inventors are key micro-level human capital for innovation (Fleming et al., 2007; Gruber et al., 2013), their mobility directly influences the number of inventions created in both departed and arrival regions (Breschi et al., 2017). In addition to inventions created by mobile inventors, scholars argue that mobile inventors facilitate knowledge transfer and spillover (Rosenkopf & Almeida, 2003; Oettl & Agrawal, 2008) and network creation, e.g., co-invention networking (Breschi & Lissoni, 2009) or firm-level alliances (Wagner & Goossen, 2018), between two departed and arrival regions.

Despite implications of the geographical mobility of inventors, causes of their geographical mobility have been less explored. Several studies suggest regional-level constraining or facilitating factors leading to geographical mobility of inventors between regions. For instance, Marx et al. (2015), from the employment policy aspect, suggests that the enforcement of non-compete agreements of a state makes inventors more likely to leave the state and migrate to other states that limit the enforcement of non-compete agreements. Taxation is also argued to be a significant dimension that causes the

[18] The results of this chapter support the validity of the instrument variable used in the study included in chapter four. Some results are included in the appendix or footnote of the study.

emigration of knowledge workers, both inventors (Akcigit et al., 2016) and scientists (Moretti & Wilson, 2017). Another study in finance suggests that the banking deregulation of the source state also drives inter-state mobility of inventors (Hombert & Matray, 2017). These studies mostly exploit regional-level regulation or policy changes and suggest factors of inventors' geographical mobility from regulatory or policy aspects, remaining individual-specific determinants underexamined.

In this study, we investigate the link between historical surname distribution and inventors' geographical mobility, thereby suggesting a demographic factor of the geographical mobility of inventors. Given the widely accepted transmission rules (Piazza et al., 1987; Rossi, 2013) and surname diversity (3,364,157 unique surnames appear in 1940 US census data), demographics on surnames are adopted in a variety of studies, such as research on migration of people, social network and mobility. Piazza et al. (1987) track migration rates of human populations using surname distribution in Italy. Degioanni and Darlu (2001) attempt to infer geographical origin of migrants in a given area using surnames. Darlu et al. (2011) show that surname distribution can be used to estimate people's mobility using the example of Savoy, France. Studies also use surnames to investigate social mobility, e.g., whether social status changes over centuries (Clark & Cummins, 2014) and whether wealth moves over generations (Clark & Cummins, 2015). In a recent study, Grilli and Allesina (2017) perform a surname analysis on academic professors to uncover and compare patterns in each academic system of the US, France, and Italy. Following these studies, we exploit inventors' surname information to test the relationship between historical surname distribution and inventors' geographical mobility. We show that the historical share of the same surname in a given location is negatively associated with the inventor's emigration from the location. The essence of our argument is that inventors are more likely to stay in a region wherein more of their families and relatives have resided.

We uncover the relationship between historical surname distribution and geographical mobility of inventors in the US. This demonstrates that inventors are less likely to emigrate a given county when a higher historic share of individuals with the same surname resided in the given county. Depending on the three specific variables of emigration, our results find that a 100 percent increase in historical surname share leads to a 0.47 percentage-point decrease, a 0.41 percentage-point decrease, a 3.2 percent decrease in the probability of leaving the county, the state, and in the emigration distance of mobility, respectively.

Several additional analyses are performed to raise the robustness of the results and suggest conditions wherein the historical surname effects are moderated or lose significance. Our results reveal that surname effect is amplified as the average value of the house owned by individuals with the same surname in the county increases, as the foreign-born ratio of individuals with the same surname in the county decreases, or when the inventor resides in a state that enforces non-compete agreements. We find no evidence that the surname effects are susceptible to invention-related inventor characteristics, such as invention productivity, quality, or length of invention experience. Results from an alternative way of testing the surname effect, i.e., using distance to geographical centroid of the surname, complement the results of the main analyses. A few placebo and robustness tests increase the confidence of the surname effect in predicting inventors' geographical mobility.

The study establishes the relationship between historical surname distribution and the geographical mobility of inventors in the US. Although the effect sizes are not large, systematic analyses of this study help us understand how the geographical mobility of inventors are, in part, determined by the historical share of the same surname in a given location. This study provides a useful and generalizable instrumental variable that can be used to correct for endogeneity in estimating the effect of the geographical mobility of inventors, thereby facilitating research on the implications of the geographical mobility of inventors.

3.2 Data and Methodology

3.2.1 Inventor data from USPTO patent data

To track inventors' geographical mobility, this study utilizes patent data from the United States Patent and Trademark Office database. Owing to the disambiguation efforts of prior researchers (e.g., Li et al., 2014), USPTO patents allow identification of individual inventors who applied and granted at least one or more patents and their geographical location with longitude and latitude information. Notably, full names are available for most inventors of granted patents, allowing us to use the surname information to calculate our focal surname variables for each inventor. Moreover, detailed information on every patent, such as application and grant dates, prior arts, technological classifications, as well as on each inventor, such as gender, are available, thereby allowing to capture various aspects of inventor individuals and have detailed investigation considering the characteristics of inventor individuals. The USPTO patent database provides such information on hundreds of thousands of inventors since 1975. Due to these merits of USPTO patent data, studies on inventor mobility often utilize the patent data to track inventor mobility and examine its causes or consequences (e.g., Hombert & Matray, 2017; Marx et al., 2015).

In this study, we use all inventors in granted patents applied between 1990 and 2010 as our initial sample. We include only inventors who have applied USPTO patents within the US for at least once, as we confined our exploratory surname variables to within-US geographical locations. We also restricted our sample to inventors with at least two patents, as we observed the mobility of an inventor using subsequent patents of the same inventor (Cappelli et al., 2019; Arts & Fleming, 2018). This leaves us with 558,227 unique inventors. We then track the geographical location histories of inventors using all granted USPTO patents applied between 1990 and 2013. Using application year of each patent and geographic location, we establish the location histories of each inventor for each year between the first and last patents applied by each inventor. As capturing the exact time that an inventor moves one location to the next location is impossible, we use the midpoint of the time window between the application years of two consecutive patents of an inventor with different geographic location information (Hombert & Matray, 2017; Marx et al., 2015). To deal with noise in location information and some prolific inventors for a given year, we use the most frequent geographic location wherein the inventor applies patents during the year and keep one observation per inventor-year-location. Finally, considering that 1940 Census data does not provide information on foreign countries, Alaska, Hawaii, and US territories in the Caribbean Sea and the Pacific Ocean, we exclude inventor-year observations that inventors are located in those locations. Thus, the sample includes only inventor-year-location observations at risk of geographical mobility and is able to measure the historical surname share based on the 1940 Census data coverage. The final sample includes 4,436,218 inventor-year observations.

3.2.2 Variables

3.2.2.1 Dependent variable

To examine geographical mobility of inventors in detail, we measure an inventor's emigration from a given county in three ways. The first and second dependent variables, *Inter-county mobility* and *Inter-state mobility*, are dichotomous variables we assign 1 if the inventor's geographical location, i.e., county and state, respectively, changes in the following year (t+1) and otherwise 0. The third dependent variable, *Distance to move-to county*, is the distance between move-from and move-to counties. This variable weighs the mobility variable using movement distance. To reduce skewness of the variable and help interpret the results, we take the natural log of *Distance to move-to county*.

3.2.2.2 Independent variable

Historical surname share, our independent variable, is the share of individuals with the same surname residing in a given county based on 1940 Census data. For instance, there were a total nine “Balsmeier”s in the 1940 Census data, and two of them resided in Sacramento county, California, which account for 22.2% of all “Balsmeier”s. Inventors whose surname is “Balsmeier” get the value 0.222 for this historical surname share variable if they reside in Sacramento county, California. For analyses, we convert the share values to percent values by multiplying one hundred and taking a natural log to reduce variable skewness and help interpret the results.

3.2.2.3 Control variables

We include several control variables at the inventor’s individual level to control for heterogeneity of inventor individual-level characteristics. *Invention experience* is measured by the number of years elapsed since the year the inventor applied for their first patent (Arts & Fleming, 2018; Conti et al., 2014). *Invention productivity* is the number of granted patents the inventor applied for between t-4 and t years (Conti et al., 2014; Hoisl, 2007). *Time spent in the county* variable is included to control for the time the inventor spent in the current county. This is calculated using the number of years after the inventor’s first arrival in the current county. *Cumulative count of mobility* variable is included to control for the geographical mobility tendency of the inventor (Hoisl, 2009). This is the cumulative count of the inter-county movements of the inventor. To control for the technological diversity of the inventor's invention portfolio, we include *Technological diversity* variable by measuring the number of unique CPC groups assigned to inventors' patents applied between t-4 and t years. *Invention quality* is the number of forward citations received during the time window between t+1 and t+3 years for patents applied between t-4 and t years (Akcigit et al., 2016; Palomeras & Melero, 2010). *Technological field* variable controls the technological fields wherein the inventor is active. As inventors with multiple inventions have multiple CPC groups in their invention portfolio, we construct a continuous variable to control for technological field heterogeneity of inventors, using the first component values from the Principle component analysis (PCA). To reflect the importance of a technological field to the inventor, we apply the term frequency-inverse document frequency (TF-IDF) weighting to all CPC groups or patent types assigned to each inventor's patents applied between t-4 and t years. We then perform PCA and take the first component values, which assign a similar value to inventors with highly related CPC groups by occurrence (El Ghaoui et al., 2013). When no CPC groups are assigned to the inventor between t-4 and t years, we use the average technological field value of the inventor.

3.2.2.4 Econometric model specification

Our baseline specification is as follows:

$$E[Y_{ijt}] = \gamma_j + \delta_t + \beta \cdot Historical\ surname\ share_{ijt} + Controls \quad (1)$$

where Y_{ijt} is the outcome variable, i.e., inter-county mobility, inter-state mobility, and distance to move-on county, in inventor i, county j, and year t. $Historical\ surname\ share_{ij}$ is the historical share of the same surname of the inventor i in the county j. γ_j denotes county fixed effects that control for time-invariant unobserved county characteristics. δ_t denotes year fixed effects that account for varying macroeconomic conditions. The parameter of interest is β. It measures the effect of the historical share of the same surname on the likelihood of inventor emigration from the location. We estimate above specification using OLS. We clustered the standard errors by the inventors to account for repeated observations of inventor individuals.

3.3 Result

Table 3-1 presents the summary statistics of the variables included in the analyses. Table 3-2 presents a set of specifications that model geographical inventor mobility. All models include year and county fixed effects. Model 1 predicts the probability of the inter-county geographical mobility. Model 2 predicts the probability of inter-state geographical mobility. Model 3 predicts the distance of geographical mobility.

Table 3-1. Summary statistics and correlations for the main variables in the analyses

	Mean	Std. Dev.	Min	Max
1. Inter-county mobility	0.050	0.219	0	1
2. Inter-state mobility	0.032	0.177	0	1
3. Log(Mobility distance)	0.307	1.407	0	8.974
4. Log(Historical surname share %)	0.216	0.544	0	4.615
5. Invention experience	7.637	7.253368	0	39
6. Invention productivity	3.340	6.549	0	935
7. Time spent in the county	4.521	4.514	0	25
8. Cumulative count of mobility	0.313	0.806	0	21
9. Technological diversity	2.815	2.864	0	213
10. Invention quality	13.701	77.079	0	57985

Note. The sample includes 4,436,218 inventor-year observations

Regressing the historical share of the same surname in a given county on geographical mobility, i.e., inter-county mobility, inter-state mobility, distance of mobility, we find a consistent pattern wherein the historical surname share decreases the probability of emigration of an inventor. Specifically, a 100 percent increase in historical surname share leads to 0.47 percentage-point, 0.41 percentage-point, and a 3.2 percent decrease in the probability of leaving the county, the state, and the distance of mobility, respectively. A 3.2 percent decrease in the distance of mobility corresponds to an absolute decrease of 2.96 miles of movement. The results are robust to a variety of control and fixed effects estimations.

Table 3-2. The effect of historical surname share on inventor emigration

	(1) Inter-county mobility	(2) Inter-state mobility	(3) Log(Mobility distance)
Log(Historical surname share %)	-0.00473***	-0.00410***	-0.0316***
	(0.000285)	(0.000306)	(0.00253)
Invention experience	-0.000461***	-0.000218***	-0.00266***
	(0.0000402)	(0.0000442)	(0.000292)
Invention productivity	0.000365***	0.0000220	0.00159***
	(0.0000628)	(0.0000412)	(0.000384)
Time spent in the county	-0.00132***	-0.000703***	-0.00769***
	(0.000155)	(0.000164)	(0.00141)
Cumulative count of mobility	0.0393***	0.0318***	0.269***
	(0.000653)	(0.000826)	(0.00474)
Technological diversity	0.00323***	0.00274***	0.0231***
	(0.000170)	(0.000181)	(0.00152)
Invention quality	-0.00000458*	-0.000009***	-0.0000616***
	(0.00000218)	(0.00000244)	(0.0000166)
Constant	0.0383***	0.0206***	0.215***
	(0.000391)	(0.000398)	(0.00275)
Technological field control	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
N.	4436167	4436167	4436154
R-squared	0.0317	0.0294	0.0346
adj. R-squared	0.0311	0.0288	0.0340

Note. Standard errors in parentheses, clustered by inventor., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models.

The results on the effects of control variables on inventor mobility are mostly consistent with general expectations. We find that inventors are less likely to emigrate from the current location as they have a longer invention experience, stayed in the county for a longer period, and higher prior invention quality. They are also more likely to emigrate from the current location as they have higher invention productivity, more diversified technological portfolio, and more frequent mobility history (e.g., Hoisl, 2009). Moreover, the effects of prior invention quality and productivity on geographical inventor mobility are found opposite to their effects on the inter-organizational inventor mobility demonstrated in Palomeras and Melero (2010). This suggests that there may exist discrepancy in the invention-related factors for geographical and inter-organizational mobility of inventors.

3.4 Additional analyses

3.4.1 Moderating effects of average value of houses

To better understand the effect of historical surname on inventor mobility, we investigate the moderating effects of house values on the surname effect. The underlying rationale is that families that owned a highly valued house would be more likely to settle around the location and, thus, enhance the surname effect on inventors' geographical mobility. In order not to be affected by the number of individuals in measuring the average value of houses, we use only household-level information and the house head's surname. Households with no house ownership information are excluded; only households with “owned” or “rented” house ownership status are included. We assign zero house value for households with a rented house. We measure the average value of houses owned by individuals (only those who are house heads in their household) with the same surname in a given county and calculate the *Average house value* variable as the logarithmic transformation of one plus the average value of houses. We estimate our specification, including *Average house value* and its interaction term with *Historical surname share* variable. Table 3-3 presents the results, and Figure 3-1 shows graphs of the interaction effects and estimated average marginal effects. Consistent with our expectations, our results show that the negative effect of the historical surname share on inventor mobility amplifies as the average house value of the surname in the given county increases.

Table 3-3. Interaction between surname share and average house value

	(1) Inter-county mobility	(2) Inter-state mobility	(3) Log(Mobility distance)
Log(Historical surname share %)	-0.00404***	-0.00373***	-0.0253***
	(0.000324)	(0.000304)	(0.00250)
Log(Avg. house value)	0.000605***	0.000899***	0.00440***
	(0.0000703)	(0.0000585)	(0.000471)
Log(Historical surname share %)	-0.000918***	-0.00106***	-0.00732***
× Log(Avg. house value)	(0.0000755)	(0.0000539)	(0.000442)
Inventor individual level controls	Yes	Yes	Yes
Technological field control	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
N.	4436167	4436167	4436154
R-squared	0.0319	0.0298	0.0348
adj. R-squared	0.0313	0.0292	0.0342

Note. Standard errors in parentheses, clustered by inventor., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models. Main variables have been mean-centered.

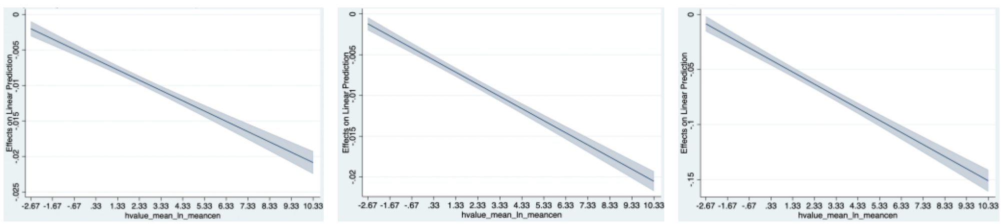


Figure 3-1. Interaction effects of surname share and average house value on inter-county mobility, inter-state mobility, mobility distance, respectively (Average marginal effects of surname share with 95% CIs)

3.4.2 Moderating effects of foreign-born ratio

We also investigate how surname effects are moderated by foreign-born ratio of individuals. As families with a higher proportion of foreign-born individuals are less likely to have settled at a location in the US, we expect that the foreign-born ratio weakens surname effects on geographical mobility. *Foreign-born ratio* variable is calculated as the logarithmic transformation of one plus the average ratio of foreign-born individuals with the same surname in a given county. We estimate our specification, including *Foreign-born ratio* and its interaction term with *Historical surname share* variable. Table 3-4 presents the results, and Figure 3-2 shows its graph of the interaction effects and estimated average marginal effects. In consistent with our expectation, the results show that the negative effect of the historical surname share on inventor mobility is amplified as the foreign-born ratio of the surname in the county decreases, that is, as more individuals with the same surname in the county were born in the US.

Table 3-4. Interaction between surname share and foreign-born ratio

	(1) Inter-county mobility	(2) Inter-state mobility	(3) Log(Mobility distance)
Log(Historical surname share %)	-0.00611***	-0.00553***	-0.0440***
	(0.000312)	(0.000360)	(0.00294)
Log(Foreign-born ratio)	0.00445**	0.00770***	0.0469***
	(0.00147)	(0.000937)	(0.00853)
Log(Historical surname share %)	0.00657***	0.00491***	0.0546***
× Log(Foreign-born ratio)	(0.00115)	(0.00104)	(0.00861)
Inventor individual level controls	Yes	Yes	Yes
Technological field control	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
N.	4436167	4436167	4436154
R-squared	0.0318	0.0295	0.0347
adj. R-squared	0.0312	0.0289	0.0341

Note. Standard errors in parentheses, clustered by inventor., * p < 0.05, ** p < 0.01, *** p < 0.001, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models. Main variables have been mean-centered.

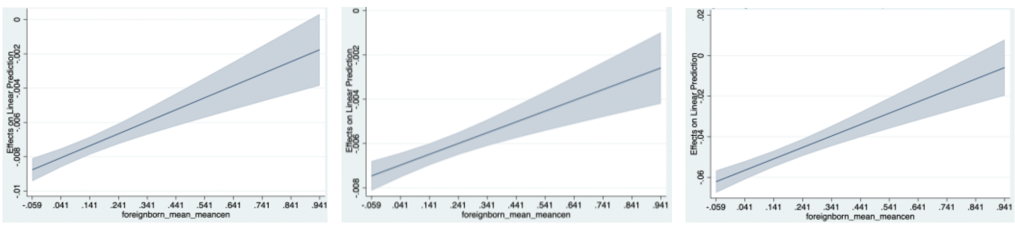


Figure 3-2. Interaction effects of surname share and foreign-born ratio on inter-county mobility, inter-state mobility, mobility distance, respectively (Average marginal effects of surname share with 95% CIs)

3.4.3 Moderating effects of inventor characteristics

We investigate the possible moderating effects of inventor characteristics on the surname effect. Specifically, we test how *Historical surname share* interact with 1) *Invention experience*, 2) *Invention quality*, and 3) *Invention productivity* to affect inventors' geographical mobility. As shown in Table 3-5, we do not find clear patterns for all the moderating effects of inventor characteristics from the analyses. Hence, we can infer that the surname effect is less likely to be susceptible to inventors' invention-related characteristics, i.e., prior invention quality, productivity, experience, for geographical mobility.

Table 3-5. Interaction between surname share and inventor characteristics (Invention experience, quality and productivity)

Panel A: Invention experience			
	(1)	(2)	(3)
	Inter-county mobility	Inter-state mobility	Log(Mobility distance)
Log(Historical surname share %)	-0.000558*	-0.000230	-0.00205
× Log(Invention quality)	(0.000220)	(0.000163)	(0.00135)
N.	4436167	4436167	4436154
R-squared	0.0318	0.0294	0.0346
adj. R-squared	0.0312	0.0288	0.0340
Panel B: Invention quality			
	(4)	(5)	(6)
	Inter-county mobility	Inter-state mobility	Log(Mobility distance)
Log(Historical surname share %)	-0.000636	0.000195	0.00108
× Log(Invention productivity)	(0.000658)	(0.000632)	(0.00551)
N.	4436167	4436167	4436154
R-squared	0.0319	0.0294	0.0346

adj. R-squared	0.0313	0.0288	0.0340

Panel C: Invention productivity			
	(7)	(8)	(9)
	Inter-county mobility	Inter-state mobility	Log(Mobility distance)
Log(Historical surname share %)	0.00103	0.000283	0.00600
× Log(Invention experience)	(0.000560)	(0.000534)	(0.00479)
N.	4436167	4436167	4436154
R-squared	0.0317	0.0294	0.0346
adj. R-squared	0.0311	0.0288	0.0340
Inventor individual level controls	Yes	Yes	Yes
Technological field control	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes

Note. Standard errors in parentheses, clustered by inventor., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models. Took natural log on Invention quality, Invention productivity, and Invention experience variables. All independent terms of Log(Historical surname share %) and invention characteristics (i.e., Invention quality, Invention productivity, and Invention experience) are included.

3.4.4 Interaction with employee non-compete agreement

Prior studies demonstrate that inventor mobility is influenced by the state enforcement of non-compete agreements on employees, which functions as a constraint for inventors' organizational mobility (Younge et al., 2015) or as a stimulus for inventors' inter-state emigration (Marx et al., 2015). As non-compete enforcement is an institutional factor directly influencing inventors' career choices, investigating how the surname effect on geographical mobility manifests depending on the enforcement of non-compete agreements is worthwhile. Most US states allow non-compete agreements, except for the following states that limit enforcement of non-competes: Alaska, California, Connecticut, Minnesota, Montana, Nevada, North Dakota, Oklahoma, Washington, West Virginia (Stuart & Sorenson, 2003; Younge et al., 2015). Therefore, we construct a dummy variable *Non-compete enforcement* with a value 1 for inventor-year observations at a county belonging to one of the states enforcing non-competes for inventors; otherwise it is assigned a value of 0. As the same non-compete clause applies for an entire state, our particular interest here is inter-state inventor emigration. Hence, among the three dependent variables of the study, we only estimate our specification on the dependent variable *Inter-state mobility*, including the *Non-compete enforcement* and its interaction term with the *Historical surname share*.

Table 3-6 presents the results, and Figure 3-5 shows its graph of the interaction effects and estimated marginal effects. The results demonstrate that the surname effect is enhanced in a state enforcing non-compete agreements. That is, when inventors reside in a state enforcing the non-compete agreements, their families or relatives become a stronger factor that attracts the inventors to stay, rather than emigrating to another state for their career. On the contrary, when inventors reside in a state limiting the non-compete agreement, the surname effect on geographical mobility is reduced.

Table 3-6. Interaction between surname share and Non-compete enforcement (DV: inter-state mobility)

	Inter-state mobility
Log(Historical surname share %)	-0.000573
	(0.00182)
Non-compete enforcement	0.00351
	(0.00224)
Log(Historical surname share %)	-0.00339†
× Non-compete enforcement	(0.00173)
Inventor individual level controls	Yes
Technological field control	Yes
Year FE	Yes
County FE	Yes
N.	4436218
R-squared	0.0253
adj. R-squared	0.0253

Note. Standard errors in parentheses, clustered by inventor. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year fixed effects included for all models. Main variables have been mean-centered.

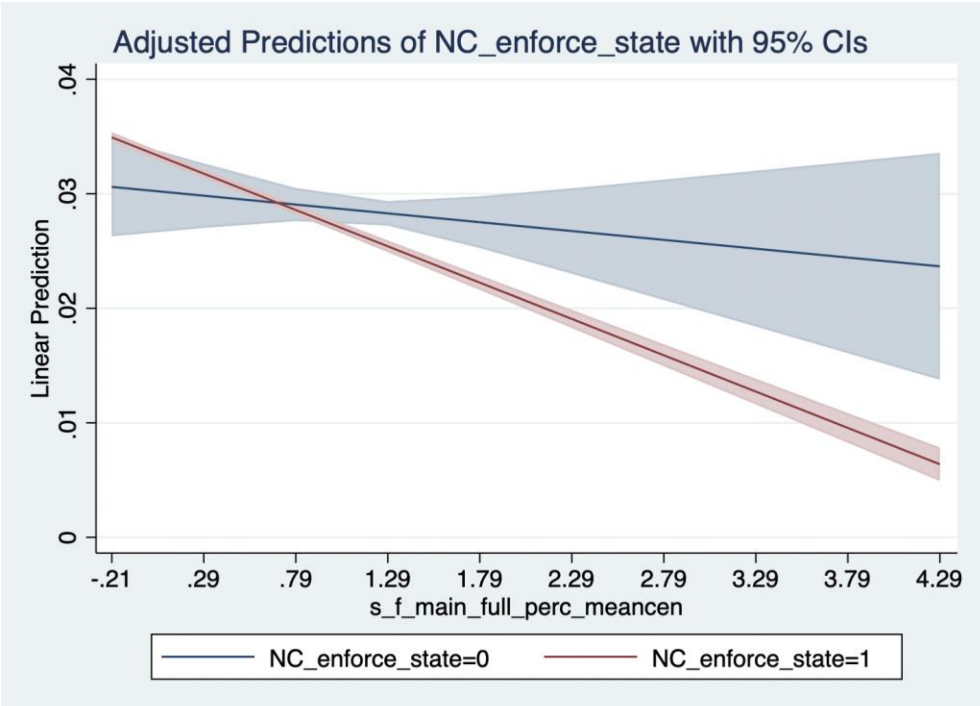


Figure 3-3. The marginal effects of surname share with 95% CIs (DV: inter-state mobility)

3.4.5 Distance-to-centroid measurement

We test the effect of historical surname distribution on inventor mobility using an alternative variable, *Distance-to-centroid*. We begin by calculating the geographical centroid of each surname using the weighted mean of the latitude and longitude of individuals with the same surname in the 1940 Census. In order to calculate this variable for a surname, it requires at least one individual with the surname resided in the US in 1940. Thus, analyzing of the surname effect using *Distance-to-centroid* variable excludes inventors with a surname that does not exist in the 1940 Census. We then measure the distance between inventor's location in a given year and the geographical centroid of their surname in 1940. We estimate our specifications to uncover how an inventor's distance to geographical centroid of their surname links to their emigration from the given location. Table 3-7 presents the results with full inventor samples. We find that a 100 percent increase in distance from a given location to the centroid of the same surname leads to 0.53 percentage-point, 0.54 percentage-point, and 4.2 percent increase in inter-county emigration, inter-state emigration, and distance of mobility, respectively.

Table 3-7. Effect of Distance-to-centroid (All inventors)

	(1) Inter-county mobility	(2) Inter-state mobility	(3) Log(Mobility distance)
Log(Distance-to-centroid)	0.00525***	0.00542***	0.0416***
	(0.000240)	(0.000281)	(0.00225)
Inventor individual level controls	Yes	Yes	Yes
Technological field control	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
N.	4059432	4059432	4059422
R-squared	0.0317	0.0307	0.0353
adj. R-squared	0.0311	0.0301	0.0346

Note. Standard errors in parentheses, clustered by inventor., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models.

3.4.6 Gender of inventors (Male vs. Female inventors)

Considering the American custom wherein women take their husband's family name upon marriage (e.g., see Goldin & Shim, 2004), we expect, generally, that the historical surname effects on geographical mobility would be significant for male inventors, rather than for female inventors. Thus, we test our specifications using two separate inventor subsamples comprising male and female inventors. Table 3-8 presents the results. Consistent with our expectations, the historical surname effect on the geographical mobility of inventors loses its significance for female inventors, while it is significant for male inventors.

Table 3-8. Effect of surname share by gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Only male inventors			Only female inventors		
	Inter-county mobility	Inter-state mobility	Log(Mobility distance)	Inter-county mobility	Inter-state mobility	Log(Mobility distance)
Log(Historical surname share %)	-0.00421***	-0.00360***	-0.0291***	-0.00314	-0.00158	-0.0159
	(0.000391)	(0.000336)	(0.00274)	(0.00174)	(0.00162)	(0.0139)
Inventor individual level controls	Yes	Yes	Yes	Yes	Yes	Yes
Technological field control	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
N.	755004	755004	755003	162516	162516	162516
R-squared	0.0315	0.0342	0.0415	0.0296	0.0230	0.0281
adj. R-squared	0.0292	0.0319	0.0392	0.0220	0.0153	0.0204

Note. Standard errors in parentheses, clustered by inventor., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Observations with singleton fixed effects are dropped before the estimation (Correia, 2015). Year and county fixed effects included for all models.

3.5 Discussion and Conclusion

The study investigates the relationship between the geographical mobility of inventors and historical surname distribution in the US in 1940. This demonstrates that the historical share of the same surname of an inventor in a given county and the geographical mobility of the inventor, in terms of inter-county mobility, inter-state mobility, and distance of mobility, are negatively associated. Our analyses show that the surname effect on geographical mobility is enhanced when the average value of houses owned by households with the same surname is higher and when the ratio of foreign-born individuals with the same surname is lower. While surname effect is valid regardless of whether the inventor stays in a state enforcing non-compete agreements, it becomes stronger in states wherein non-compete agreements are enforced. Our analyses do not find significant moderating effects of invention-related characteristics for the surname effect on the geographical mobility of inventors. An alternative way of estimating the surname effect, i.e., estimation with the distance-to-centroid variable, provides consistent results, complementing the main results found with historical surname share variable.

Results from some additional tests show the more detailed conditions wherein the surname effect becomes more effective in predicting the geographical mobility of inventors. Specifically, while the surname effect is significant for male inventors, it is not significant for female inventors. Thus, these criteria should be considered when using the surname variable to track inventors' geographical mobility in future studies.

Our study contributes to the research stream on geographical mobility of inventors by providing a useful and generalizable instrumental variable for the geographical mobility of US inventors. Although prior studies emphasize the important role of inventors moving across regions (Almeida & Kogut, 1999; Jaffe et al., 2000; Rosenkopf & Almeida, 2003), systematic and direct investigation of the mobility effect and its size has often been limited owing to the endogenous nature of the geographical mobility of inventors and the observational data used to track their mobility. Inventor mobility decisions are endogenous to empirical models of various pre- and post-mobility factors, such as performance (e.g., invention productivity and quality) and invention-related behaviors (e.g., collaboration). Additionally, designing a randomized experiment for inventor mobility study is difficult (and almost impossible). That is, instructing the geographical mobility of inventors and randomly treating a certain inventor or inventor group to investigate the implications of inventor mobility is difficult (or too expensive). Moreover, as patent data allow for identifying a large number of inventors and tracking their locations (Li et al., 2014), researchers in recent studies rely on patent publication data to investigate inventor mobility and its consequences. For these reasons, models estimating the effects of the geographical mobility of inventors often confront causal ambiguity, errors-in-variables, or omitted variable problems (Bascle, 2008). Without controlling for potential endogeneity, the estimation of inventor mobility effects would become biased, and the causal inference would not be accurate.

By using historical surname distribution in the US and uncovering its effect on the geographical mobility of inventors in the US, the study provides an instrumental variable for inventors' geographical mobility. As USPTO patents and the inventor disambiguation data provide surnames of most inventors with at least one patent applied to USPTO, these historical surname variables, i.e., historical surname share and distance to geographical centroid variables, are generalizable for all US inventors. Furthermore, this surname variable can be aggregated to the level of organization or region to instrument for the effect of inventor mobility at the organizational or regional level.

The results of the interaction analyses increase the credibility of the historical surname variable as an instrumental variable for the geographical mobility of inventors. Specifically, the moderating effects of the average house values of households with the same surname in a given county and the ratio of foreign-born individuals with the same surname in a given county on the surname effect are found to be significant. This raises confidence that the geographical mobility of inventors is influenced by historical demographics, although this might be a minor factor. Moreover, no consistent pattern found in the analyses on the moderation of invention-related characteristics suggests that the surname effect is less likely to be susceptible to other invention-related characteristics of the inventor. Therefore, this increases the credibility of using historical surname variables to instrument the geographical mobility of an inventor when estimating the mobility effects on post-mobility invention-related outcomes.

The study also contributes to the research stream focusing on non-compete enforcement and its influence on inventor mobility. Prior studies suggest that the non-compete enforcement of a state affects highly skilled knowledge workers (Stuart &

Sorenson, 2003). Focusing particularly on inventors, studies suggest that non-compete enforcement influences inventors' inter-organizational mobility (Younge et al., 2015), inter-state mobility (Marx et al., 2015), and inter-technological field mobility (Arts & Fleming, 2018). Extending this research stream, this study examines how historical demographics, i.e., historical surname share, interact with non-compete enforcement to influence inter-state inventor mobility. Our results show that the surname effect on inter-state mobility is amplified in a state wherein non-compete agreements are enforced. When inventors' career choices are constrained by non-compete agreements, they are more likely to choose to stay around their families or relatives.

These findings on the interaction effect between non-compete enforcement and historical surname share provide a few implications for firm managers and policymakers. Firms located in a state enforcing non-compete agreements may place inventors in a regional location with more of their families or relatives to prevent their departure as the surname effect is more effective in these states. However, this study suggests that the surname effect becomes weaker in states limiting enforcement of non-compete agreements. Thus, firms in such states should devise other measures to keep their inventors as not only institutional constraints but also surname constraints of geographical mobility are not as strong in such states, compared to states enforcing non-compete agreements. For policymakers, the results suggest that enforcing non-compete agreements allows to retain inventors within the state and prevent departure of inventors to other states, especially when the state was historically comprised of a large population of individuals with the same surname of inventors currently working in the location.

Chapter 4. Inventor mobility and entrepreneurial ecosystem[19]

4.1 Introduction

Entrepreneurship—especially when driven by novel technologies—has been recognized as an essential source of economic growth and improved quality of life since Smith (1776) and Schumpeter (1942). Recent evidence confirms that newly-founded firms are responsible for job creation (Decker et al., 2014; Glaeser et al., 2015), productivity (Gennaioli et al, 2013), and additional innovation (Kortum & Lerner, 2001; Lee, 2013). Unsurprisingly, policymakers worldwide have sought to spur startup activity, often in hopes of replicating the entrepreneurial dynamics of California's Silicon Valley. That so many efforts have fallen far short (Lerner, 2009) speaks to a lack of understanding and causal evidence for an earlier stage in the chain: if entrepreneurship drives economic growth, what drives entrepreneurship? Further, given that the vast majority of new firms fail (Haltiwanger et al., 2013)—including 75% of venture-capital backed firms (Hall & Woodward, 2010)—what are the critical inputs for successful startups?

The co-occurrence of the words "innovation and entrepreneurship" is ubiquitous in both academic and popular circles. In this paper, we examine whether productive entrepreneurship (i.e., successful startups) depends critically on innovation—or, more

[19] This chapter is adapted from joint work with Benjamin Balsmeier, Lee Fleming, and Matt Marx. For a working paper version, please see: https://doi.org/10.3386/w27605

precisely, on the inventors who are responsible for innovations.[20] To be sure, scholars have long observed that human capital, including technical talent, is an important ingredient in the entrepreneurial recipe. Lerner & Nanda (2020) claim that "[r]egions like Silicon Valley have an abundance of resources for entrepreneurs, [including] excellent engineers…" Jensen & Thursby (2001) likewise argue that scientific inventors need to be fully engaged and motivated for technologies to be successfully commercialized in new firms (see also Zucker et al, 1998; Marx & Hsu, 2021).[21] Larger-scale, if suggestive, evidence for the role of inventors in high-growth entrepreneurship comes from correlations between the supply of technical workers' levels of patenting, entrepreneurial firm founding, and employment (e.g. Kerr, 2013; Maloney & Caicedo, 2016; Azoulay et al., 2020). Glaeser & Kerr (2009) find that talent explains 60-80% of the variance in regional entrepreneurship in U.S. manufacturing, concluding that "the broad stability of this finding suggests that people and their human capital are probably the crucial ingredient for most new entrepreneurs" (p. 659).

Indeed, even absent causal evidence it might seem self-evident that inventors play an essential role in high-growth entrepreneurship. Steve Wozniak, who invented what would

---

[20] Our focus on inventors builds on recent advances in disambiguation, which enabled identification of these inventors and facilitated progress on the question of how individuals contribute to innovation, productivity, and economic growth (Bhaskarabhatla et al., 2021; Kline et al. 2019; Azoulay et al. 2020).

[21] Not all high-growth firms in the U.S. are high-tech, and vice-versa. However, Hathaway (2018) reports that high-tech firms are overrepresented by 4x among high-growth firms (21% vs. 5% of all firms) as defined by *Inc*. Magazine's annual list of the 5,000 fastest-growing privately held firms in the U.S (see also Lerner and Nanda, 2020).

become the Apple I while working at Hewlett-Packard, famously could not convince his superiors to commercialize the invention and subsequently left to found a new firm with Steve Jobs. At the same time, several successful startups including Slack, Skype, Whatsapp, Alibaba, and BaseCamp, largely contracted out engineering and product development activities to geographically distant locations, and investors regularly pressure their portfolio companies to outsource technical development. As Jim Breyer, managing general partner of Accel Partners, remarked: “There isn't a board meeting that goes by that we don't ask, Why aren't you being more aggressive [with software development] in India and China?"[22] Therefore, the direction of causality between the supply of technical talent and entrepreneurship remains unclear (Burchardi et al., 2020). The correlations found by Glaeser & Kerr (2009) could reflect not a causal effect of talent on entrepreneurship but rather the flocking of skilled workers to opportunity. Or, it might be that investors like Jim Breyer are correct and technical talent is simply not as important as conventional wisdom might (like to) assume.

In pursuit of causal evidence on this point, we investigate how the supply of key technical talent—including technology- and task-specific capital (Gibbons & Waldman, 2004)—influences the funding and success of high-growth ventures.[23] We focus on

---

[22] https://www.sfgate.com/business/article/looking-offshore-investors-vc-firms-push-for-2813526.php

[23] Related to this paper, several studies have addressed the role of local inventors in regional productivity. For example, Agrawal et al. (2011) show that inventor emigration decreases local knowledge flow in the source region but also drives knowledge back into the departed region. A growing and influential literature on foreign immigration suggests positive impacts on the U.S. of an influx of inventors from outside its borders, including greater patenting and innovation (Bernstein et al., 2018; Hunt & Gauthier-Loiselle, 2010; Burchardi

venture-backed startups as a particularly promising subset of new firms. Although only 0.5% of new businesses obtain venture financing (Puri & Zarutskie, 2009), nearly one-half of firms that complete an Initial Public Offering had received venture capital backing (Lerner & Nanda, 2020). We address reverse-causality concerns by instrumenting inventor inflows with the share of inventors' surnames in a county based on the nationwide distribution of surnames from the 1940 U.S. Census. Our shift-share instrument represents an advance over prior efforts in two ways. First, because the "shares" stem from more than three million unique surnames across more than 3,000 counties, it is less vulnerable to critiques of such instruments with low variation or a few highly-determinative shares (see Goldsmith-Pinkham et al., 2020, Adao et al., 2019, and Borusayak et al., 2018, for a fuller discussion of the issue). Second, focusing on the U.S. lessens concerns regarding endogenous origin-destination combinations (e.g., Indian engineers migrating to Silicon Valley) and also addresses the issue of potential endogenous choice of regions and selection of incoming inventors at the national level (Moser et al., 2014; Parey et al., 2017).

We find that the (exogenous) arrival of inventors in a county has a substantial impact on entrepreneurial activity. Arriving inventors increase the number of venture-backed startups in a county, in the same sectors as the arriving inventors and at the expense of other sectors. Not only does the arrival of inventors produce more startups; the influx of

---

et al., 2020; Kerr & Lincoln, 2010), wages (Peri et al., 2015) and TFP (Capelli et al., 2019). Our study differs from these in that we study internal migration and entrepreneurship.

technical talent yields startups with more successful outcomes (IPO or attractive acquisition). Our preferred empirical specifications suggest that counties may expect one additional venture-backed startup for every 28 additional inventors, whereas a successful startup requires an additional 460 inventors. Incoming inventors even contribute to an increase in the number of “unicorn” startups (i.e. exit valuation exceeding $1B). However, the increase in successful exits is not merely the result of more “shots on goal”; these correspond with a reduction in bankruptcies as well as so-called “fire-sale” acquisitions. Therefore, the local availability of technical talent appears to improve the efficiency of venture investment, reallocating away from failed, low-tech startups. These results are robust to a variety of alternative instrument specifications and placebo tests and are moreover not restricted to the top ten counties by entrepreneurial activity (Silicon Valley, etc.).

4.2 Data

We assemble three different sources at varying degrees of aggregation and times to arrive at a panel dataset at the county-year level.

4.2.1 Historic Census data

We begin with the complete 1940 U.S. Census records for 131,940,709 citizens in 38,382,088 households (http://sites.mnhs.org/library/content/1940-census). As we will explain in detail in the next section, our identification strategy relies on being able to observe the name and location of each U.S. citizen in 1940 in order to predict inventor moves. The historic data include 3,363,932 different surnames, of which 27% appear only once. (The median is 3, mean is 39, and maximum is 1,359,079 for Smith.) Figure 4-1 illustrates the sparse geographical distributions of “Marx”. After some cleaning and standardizing procedures, described in detail in Appendix 2, there were 42,268 Flemings, 6,232 Marxes, 153 Shins, and 9 Balsmeiers in the 1940 census data. All analyses below are robust to excluding prolific surnames as indicated by high (local) frequency or wealth, e.g. the Smiths and Rockefellers. The 1940 U.S. Census records consist of 3097 counties and other districts based on the county system in 1940. In order to help matching with the location information of inventors, we translate 19 counties or districts, which are old and no longer in use, to the 2020 concordance. (based on https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html from 1970 to 2020). Please see Appendix 2 for details on geographic disambiguation.

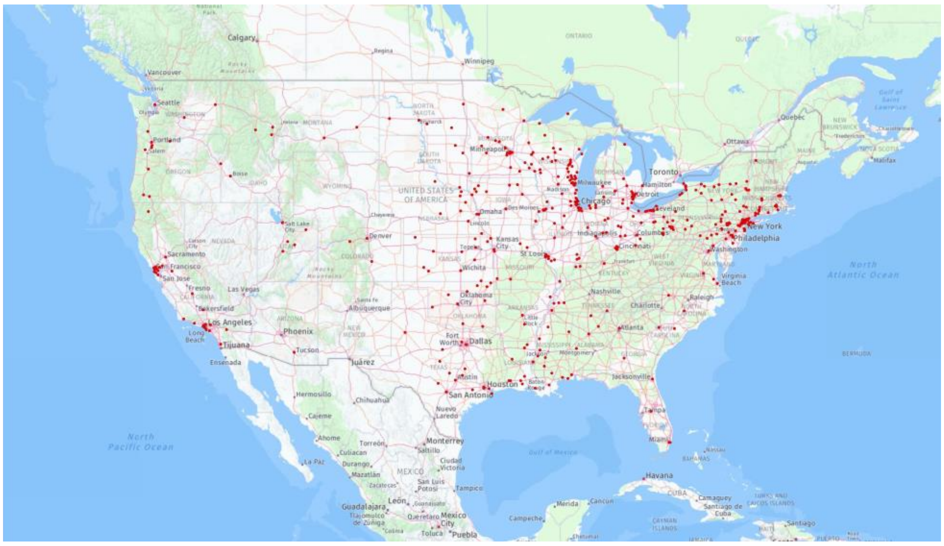


Figure 4-1. Spatial distribution of the surname “Marx” in 1940 (each red dot = 50

individuals).

4.2.2 Inventor data

We begin with raw data from the United States Patent and Trademark Office (USPTO) from 1976-2018 (only the intersection of patent and entrepreneurship data are used, see below). Although the USPTO lists inventors for every patent, it does not provide unique identifiers for them. For example, even the relatively rare name of Matthew Marx is listed as inventing many patents, including 5,995,928, "Method and apparatus for continuous spelling speech recognition with early identification, 6,173,266, "System and method for developing interactive speech applications," and 7,271,262, "Pyrrolopyrimidine derivatives." In this simple example, it would seem reasonable based on the titles alone that the same inventor authored the first two but not the last patent, and that is indeed the case. Inventor names can be disambiguated with a variety of algorithms, here we use Balsmeier et al. (2018). After applying the name cleaning and standardizing procedures and the matching algorithm, described in detail in Appendix 2, we match 91.1% of inventors' surname to a surname from the 1940 Census. Note that the name cleaning exercise has no significant effect on the size of the estimated coefficients but decreases matching errors and improves precision of the instrument.

We used the inventor ID and location to identify inventor moves across U.S. counties. We drop all inventors with a single patent. Then, using patent application year as a timestamp, we count an inward move in the first year we first observe an inventor in a county. As noted by Cheyre et al. (2015), patent application dates do not necessarily correspond with dates of employment and in particular may lag actual moves. Hence, the inventor may have moved into a county earlier than we detect, leading to a fuzzy lower bound of the actual lag between our variable of interest and the actual inward moves. In 96% of cases, we observe an incoming inventor patenting elsewhere within 5 years earlier (mean = 2.6). Results are robust to excluding inventor moves with longer gaps between two patenting events, or temporary stops at a third county. If an inventor appears on two or more patents within a given year, we follow Moretti & Wilson (2014) and take the most frequent location.

4.2.3 Entrepreneurship data

To measure high-growth entrepreneurship, we use VentureXpert, which is part of Thompson's economic data suite and covers all venture-backed firms in U.S. It offers detailed information on the location, industry classification and significant growth events (M&As and IPOs) of the funded companies. The data is sourced from venture capital firms, company filings and various news sources.

Our baseline sample consists of all startups with available information on founding year, industry and location, starting in 1987 as VentureExpert lacks comprehensive coverage beforehand. Our sample ends in 2007 to avoid truncated measures of whether a startup achieved a significant event (successful M&A or IPO) within ten years since founding. It is worthwhile to note that our sample of venture-backed startups represents a positive selection of startups as VCs typically only fund firms with attractive growth prospects. We focus on such events because they drive economic dynamism, innovation, and long-term economic (Lerner & Nanda, 2020). Although this focus ignores other types of entrepreneurship, e.g. hairdressers, nail polish studios and various sole proprietorships, where productivity growth has been notoriously difficult to achieve because of limited possibilities to leverage technological progress (Baumol & Bowen, 1966), our sample retains low-technology VC backed startups. Figures 4-2, 3, 4, 5 illustrate that the local supply of inventors and high-growth entrepreneurship are indeed strongly correlated, follow similar trends over time, and tend to be regionally clustered.

While suggestive, Figures 4-2, 3, 4, 5 cannot speak to whether these patterns reflect self-selection of inventors towards previously successful regions or whether there exists an arguably causal link between the local supply of inventors and entrepreneurship. Furthermore, they fail to differentiate between successful and failed startups and between high-tech (biotechnology, life science, computer and communication and semiconductor industries) and low-tech startups (various categories ranging from food processing to transportation as explicitly defined by VentureXpert). Figures 4-6 illustrates the spatial distributions of these technological categories and mobile inventors and implies a technology-specific link between inventors and startups that can be exploited both theoretically and econometrically.

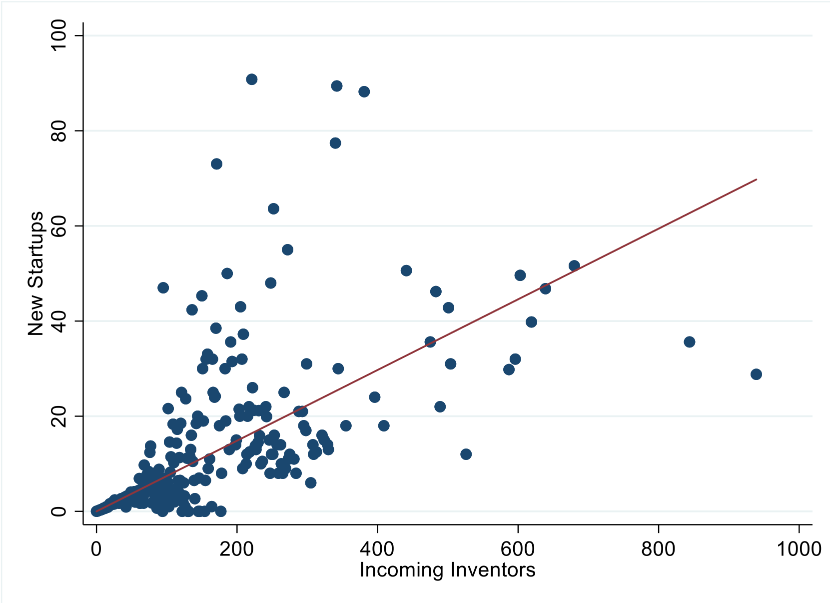


Figure 4-2. Graphical representation of raw data

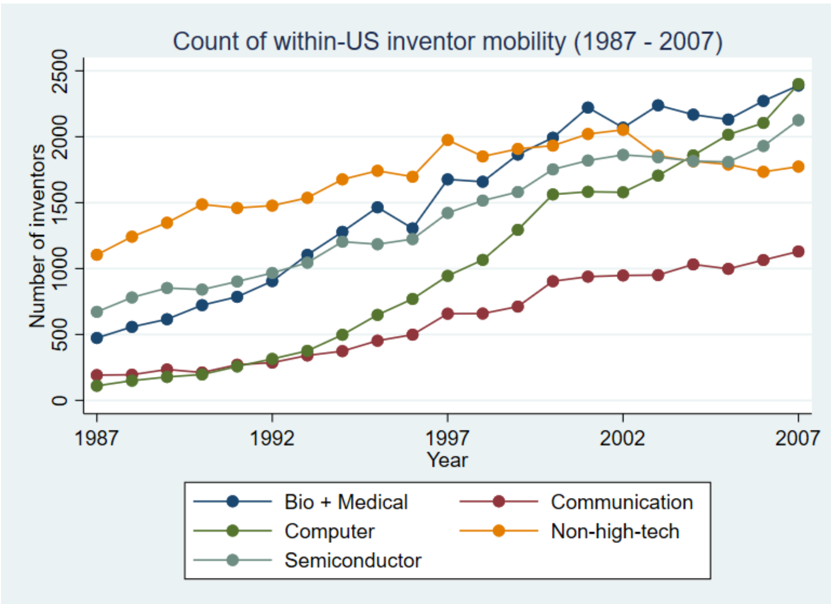


Figure 4-3. Graphical representation of US inventor moves

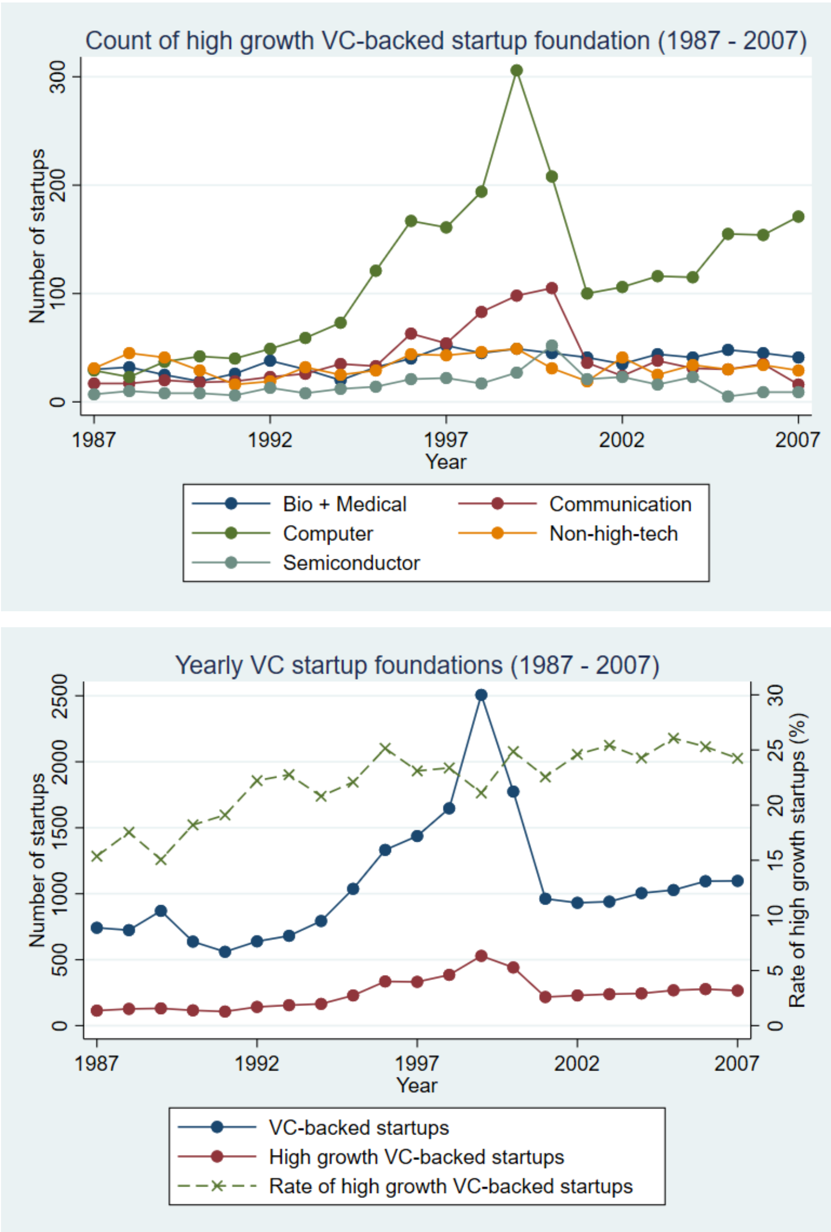


Figure 4-4. Graphical representation of venture-backed startup creation

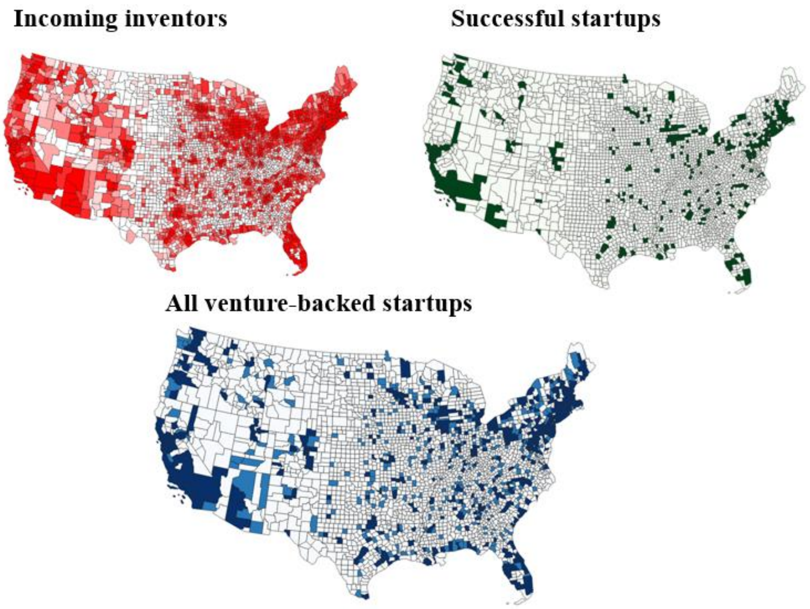


Figure 4-5. Geographical clustering of inventor moves, startups, and high-growth startups, 1987 to 2007

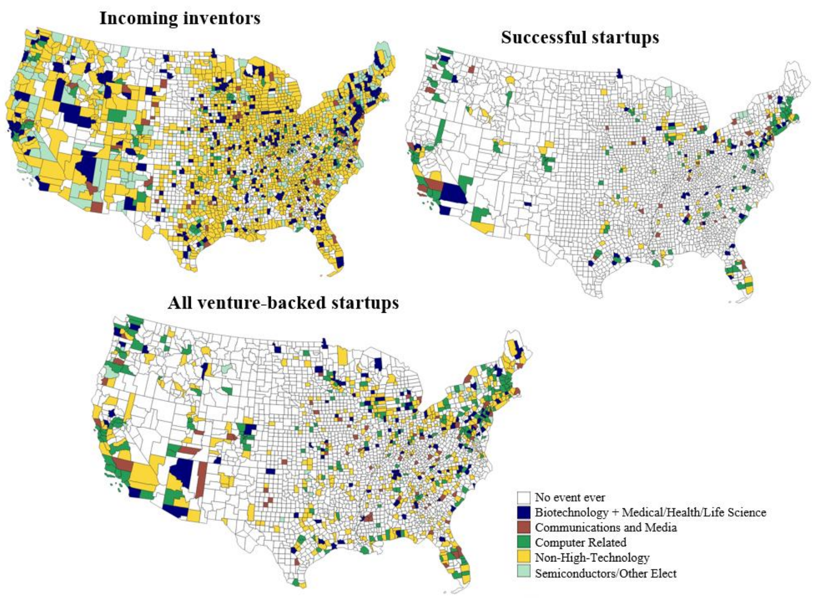


Figure 4-6. Geographical clustering of inventor moves, startups, and high-growth startups, 1987 to 2007 by major technology within county

Following Ewens & Marx (2018), we define a successful startup as having completed a merger, acquisition, or initial public offering with valuation exceeding 125% of the total invested venture capital within 10 years since founding. We also measure a 500% return on invested capital as well as “unicorns” i.e., startups which exit with a valuation of $1B or greater. Because VentureXpert is missing many acquisition values (and some IPO values), we fill these in using data from Pitchbook and Crunchbase via exact match on website URL and state (Dorn et al., 2020). Where VentureXpert was missing capital invested, we filled in these values from those databases in order to calculate the return on invested capital. We also filled in founding years from the databases when they are missing in VentureXpert.

For failed startups, we used the current status of each startup indicated in VentureXpert, indicating failure if they were listed as “Defunct” or “Bankruptcy.” Differentiating between failed and successful startups is crucial, as Decker at al. (2014) show that it is the few high-growth startups that survive the first ten years of their existence that are responsible for about 50% of US gross job creation. Where a startup had not exited within ten years of founding, it was neither counted either as having failed or succeeded.[24]

---

[24] Considering sectors where startups often take longer than 10 years to make it to an exit such as biopharmaceutical industry, we also test with a 12-year window, instead of 10-year window, to capture the successful and failed startups. We use only county-year observations between 1987 and 2005 as we cut the last 2 years to allow for 12 years of observations. The estimation results are almost identical to our main

Table 4-1 provides descriptive statistics of the dataset at the county-year level. For the sample of 27,619 venture-backed startups, 26% achieve an M&A or IPO within 10 years of their foundation, with an average return of 1646% (median 203%) on the invested capital (note that these values may be over-estimated as we calculated based on only ones that their exit values are available). For failed startups in our sample, there are total 3386 venture-backed startups that are indicated as "Defunct" or "Bankruptcy" in VentureXpert.[25]

Table 4-1. Descriptive statistics at U.S. county level, N=65,247

Variable	mean	median	std dev	min	max
Number of incoming inventors	2.15	0.00	11.96	0.00	700.00
Instrument	1.98	0.70	8.68	0.00	356.03
Number of overall venture-backed startups	0.42	0.00	4.20	0.00	314.00
Number of successful startups (RoR $\geq$ 125%)	0.04	0.00	0.60	0.00	38.00
Number of successful startups (RoR $\geq$ 500%)	0.02	0.00	0.28	0.00	19.00
Number of successful startups (Exit $\geq$ 1B)	0.00	0.00	0.05	0.00	4.00
Number of failed startups	0.05	0.00	0.83	0.00	91.00
Number of failed startups (inc. RoR < 125%)	0.08	0.00	0.08	0.00	123.00
Number of high-tech startups	0.34	0.00	3.79	0.00	306.00
Number of low-tech startups	0.09	0.00	0.70	0.00	38.00

Notes. This table reports summary statistics of the key variables used in our regression analyses at the county level, covering 3107 counties 1987-2007. "Successful" startups are those that complete either an IPO or successful acquisition within 10 years, and we have three different cutoffs at an exit value $\geq$ 125%, 500% of total venture capital acquired or an absolute exit value $\geq$ 1B dollars. "Failed" startups are those that are currently "Defunct" or "Bankruptcy" as indicated in

results.

[25] In unreported naïve regressions of local GDP growth on successful and failed startups we find a significant and positive effect of successful startups as opposed to an insignificant effect of failed startups.

VentureXpert database. In addition, we have another variable for "Failed" startups that includes startups that complete either an IPO or successful acquisition within 10 years, but achieve a value < 125% of total venture capital acquired. High- vs. low-tech startups are categorized according to VentureXpert classifications.

4.3 Shift-share instrument construction

We want to find the impact of inventor inflows on entrepreneurship. We can estimate this via OLS:

$$Y_{d,t} = \alpha_0 + \beta \cdot Inv_{d,t-1} + \delta_t \times \eta_d + \gamma_d + \varepsilon_{dt} \tag{1}$$

where $Y_{d,t}$ is a dependent variable observed for county d in year t. $Inv_{d,t-1}$ is the number of inventors who moved to county d in year t-1. δ_t denotes year fixed effects and η_d denotes state fixed effects. We control for state-year specific shocks, such as varying state-level economic conditions and policy changes, through state-year fixed effects $\delta_t \times \eta_d$. γ_d controls for time-invariant unobserved county characteristics that may confound our identification of β. ε_{dt} is the error term.

The key econometric challenge with Equation (1) is that unobserved factors influence both the rate of incoming inventors and local economic conditions; for example, innovative counties are attractive to inventors. Although county fixed effects will effectively control for any persistent differences in innovation levels across counties, this misses temporary local trends that might attract inventors. To address this threat to identification, we construct a shift-share instrument for inventor inflows that builds on the work of Bartik (1991) and its application to international immigration to the U.S. (Card, 2001). Prior studies had noted that immigrants tend to locate near previous immigrants from the same country of origin (Bartel, 1989; Lalonde & Topel, 1991). Card (2001) and others (see Jaeger et al. (2018) for an overview) exploited this observation to predict immigrant inflows into particular regions, by interacting past shares of immigrants from an origin country to a given region with the contemporaneous total inflow or shift of migrants from the same country at the national level.

We leverage this idea to create an instrument for the contemporaneous inflow of U.S. inventors to a certain county based on the spatial distribution of U.S. surnames across counties in 1940. The intuition is simple: although a host of factors influence where inventors locate—or, more important to our study, re-locate—on the margin, an inventor should prefer to move to a county where there are likely to be more relatives. Although we lack data on family structure and relationships for the entire population of U.S. inventors, we borrow an approach from the immigration literature which utilizes the observation that people with a certain family name are found more frequently at places where there were other people with same name in the past (see Darlu et al. (2011) for the example of Savoy, France and Clark & Cummins (2015) for England). We illustrate below that these patterns hold for individual US inventors. Specifically, we define our instrument as:

$$\widetilde{Inv_{dt}} = \sum_n \frac{P_{dn}^{1940}}{P_n^{1940}} \cdot Inv_{nt} \quad (2)$$

where P_{dn}^{1940} is the population of people in county d with surname n in 1940, P_{n}^{1940} is the number of people with surname n in the entire U.S. in 1940 and Inv_{nt} is the number of inventors with surname n who move from any county in the U.S. to any other county in the U.S. in year t. The expected inflow of inventors $\widetilde{Inv_{dt}}$ in county d at time t is thus the weighted sum of inventors that move across the U.S. with surname n (the "shift") with the historical distribution of the same family names (the "shares") serving as weights. The intuitive appeal behind this instrument (as in prior immigration studies) is that it generates variation at the local level by exploiting variation at the national level, which is arguably not influenced by local conditions. (That is, the total number of inventors with the name Fleming who move from within the entire U.S. is unlikely to be driven by the local economic conditions of one out of the more than 3,000 U.S. counties.)

4.3.1 Variation and non-persistence of the county-level instrument

One advantage of this instrument over prior shift-share instruments generally, and settlement instruments in particular, is the greater variation in the distribution of names (i.e., the "shares") that stem from more than 3 million unique surnames in 1940 across varying destination and origin areas. (By contrast, immigration studies typically analyze 192 different countries, often with particularly influential origin-destination relationships.) Our estimation should therefore be less vulnerable to problems that arise from low independent variation in shares or overly strong influences of a single or few shares (see critiques in the recent literature, Borusyak et al., (2018), Goldsmith-Pinkham et al. (2020), or Adao et al. (2019), and our placebo tests below).

A second advantage of our U.S.-focused shift-share instrument is that a given surname is typically not bound to a specific county of origin (as is more common with country-level analysis, see Moser et al. (2014); Parey et al. (2017)). Thus, the spatial distribution of the origin of mobile inventors with a surname varies substantially over time (and is the only variation we exploit in our IV). This makes an endogenous origin-destination combination (such as Indian engineers coming into Silicon Valley over long periods of time) highly unlikely to drive our results. Put differently, that mobile inventors with certain names come from various origin counties means that it is less likely that our "shift" is correlated with unobserved endogenous characteristics of origin areas. The considerable variation in the distribution of surnames over time also addresses the "persistence problem" with shift-share instruments in the immigration literature (Jaeger et al., 2018). Our instrument thus minimizes serial correlation between specific origin and destination regions, as criticized in studies of international migration.

Figures 4-7, 8, 9 illustrate the variation over time and space with the example of all inventors that moved across the U.S. between 1976 to 2015 and have the last name Fleming (75 moves in total). Figure 4-7 shows how the number of mobile inventors with surname Fleming varies over time yet does not exhibit a trend. The maps in Figure 4-8 show to which counties the Flemings moved to in the 1980s, 1990s, and 2000s, and the maps in Figure 4-9 show the origin counties the Flemings moved away from in the 1980s, 1990s, and 2000s, illustrating significant variation in origin and destination counties over time. In fact, to our eyes only one or two of the counties from which Flemings emigrated in the 1990s was also a significant source of Flemings in the 2000s (Figure 4-9). The same appears true for destination counties in Figure 8, as just one example of why our county-level instrument should be less susceptible than a country-level instrument to the "persistence problem."

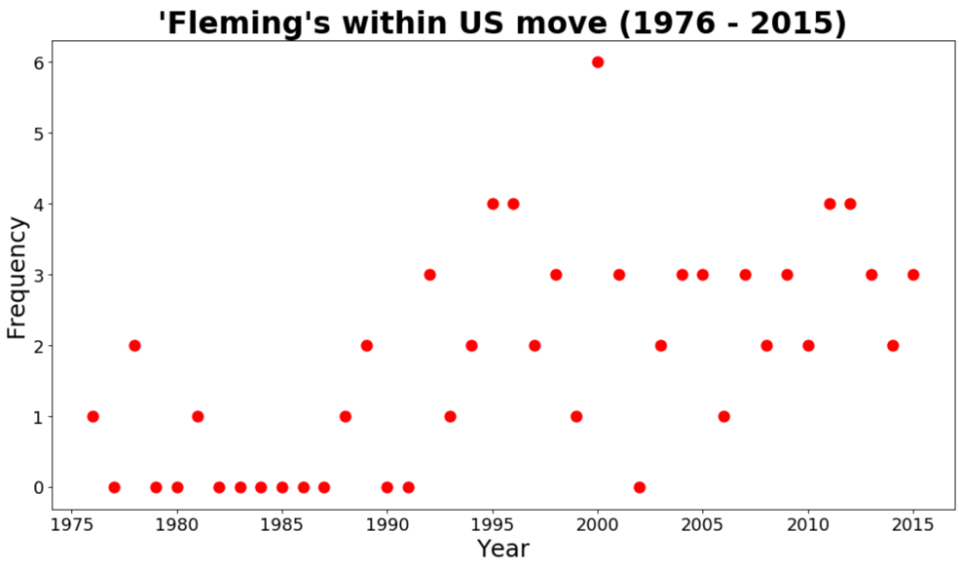


Figure 4-7. Frequency of moving inventors within the U.S. named Fleming over time

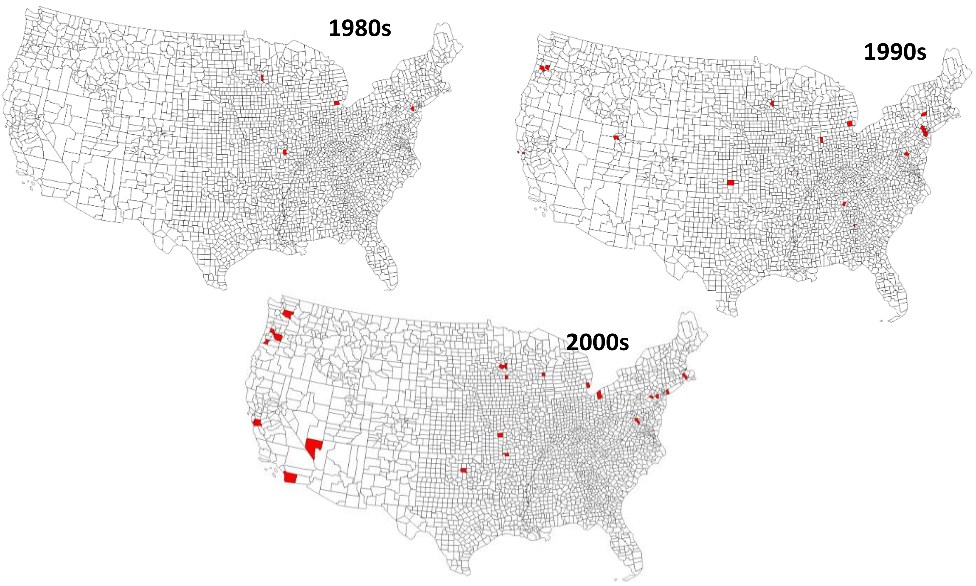


Figure 4-8. Destination counties of moving inventors within the U.S. named Fleming

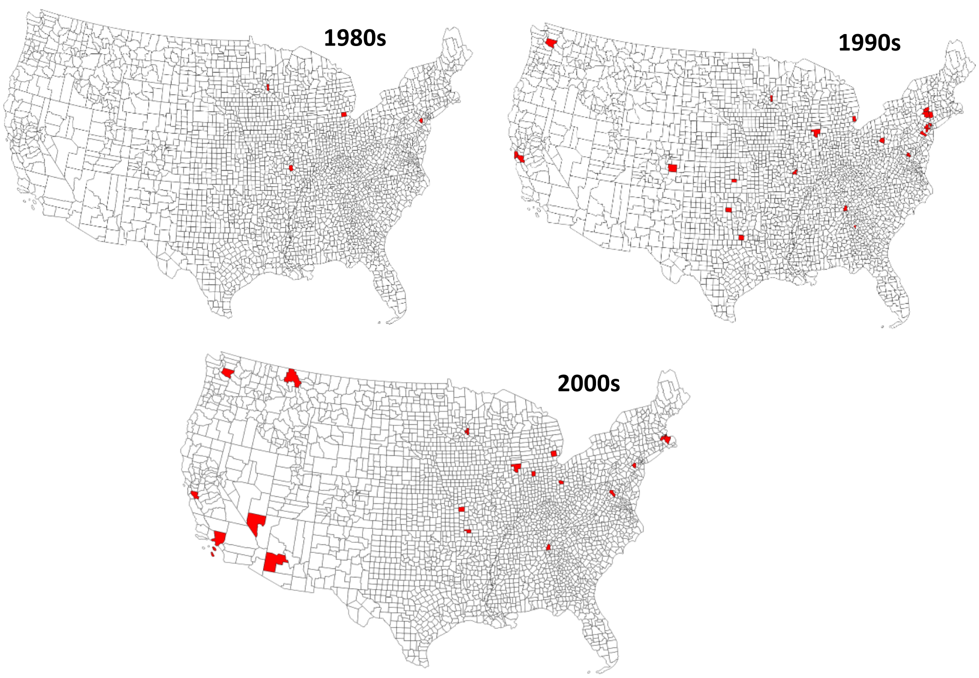


Figure 4-9. Origin counties of moving inventors within the U.S. named Fleming

4.3.2 Final instrument with "leave-out"

A remaining concern could be that at least some national movements of inventors are still driven by local economic conditions, and that these might be correlated with past shocks. It could be, for instance, that inventors and families with the name Fleming were always interested in mechanical engineering and thus would have settled in areas where mechanical engineering was in high demand in 1940. If the same area experiences a high demand in mechanical engineering today, then inventors with the name Fleming might be more likely move to that region for endogenous reasons. To reduce these endogeneity concerns, we leave out county d's own inflows from the national flow of inventors with the same surname (see Buchardi et al. (2020), Wozniak & Murray (2012), or Hunt (2017) for similar approaches). Our preferred instrument is thus:

$$\widetilde{Inv_{d,t-1,leave-out}} = \sum_n \frac{P_{dn}^{1940}}{P_n^{1940}} \cdot Inv_{n,t-1,leave-out(n,d)} \tag{3}$$

where $Inv_{nt,leave-out(n,d)}$ is the total number of inventors with name n who move to counties outside of d. The leave-out strategy ensures that the potentially-endogenous choice of Flemings to move to county d does not drive changes in our instrument.

It should be noted that both stages of our IV include county fixed effects. Identification thus derives from weighted time-varying changes in the number of moving inventors for a given surname at the national level, excluding those moving to county d, combined with representation of the same surname in county d in 1940. Ideally,

$Inv_{dt,leave-out}$ is truly exogenous and can be used to estimate the causal impact of inventor inflows on using Equation (1), instrumenting $Inv_{d,t}$ with $Inv_{dt,leave-out}$ as in (3).

4.3.3 Industry-specific version of the instrument

Although most of our analyses focus on the overall number of inventors who move into a certain county, we are also interested in the industry-specific inflow of inventors and their influence on the industry-specific rate of startup foundation, e.g. how many biotech startups are founded in a county in response to the inflow of inventors with a biotech background. To this end, we create an additional dataset at the destination county-industry-year level. We differentiate between each of the four high-tech classifications and the low-tech sector as defined by VentureXpert. We match inventors with these industries based on the technology classification assigned to each patent. If an inventor filed patents in more than one tech class we used the most frequent and in case of a tie the earliest (see Appendix 2 for details). Armed with this dataset we can, similar to (1), estimate the following equation with OLS:

$$Y_{d,i,t} = \alpha_0 + \beta \cdot Inv_{d,i,t-1} + \gamma_d \times \delta_t + \theta_i \times \delta_t + \theta_i \times \gamma_d + \varepsilon_{dt} \tag{4}$$

where $Y_{d,i,t}$ stands for a dependent variable observed for county d, industry i at time t. $Inv_{d,i,t}$ is the number of inventors with a technological background closely related to industry i that moved to county d in year t-1. The key difference to (1) is that we can control for county-year specific shocks through county-year fixed effects $\gamma_d \times \delta_t$, i.e. we can effectively control for any unobserved county characteristic, irrespective of whether it is varying or not varying over time. This includes, for instance, the total number of inventors. Put differently, identification of β will only come from relative differences across industries within a county and year. Hence, we only expect β to be positive if, for instance, a higher fraction of biotech inventors out of all inventors moving into a given region at a given time would lead to a higher fraction of biotech startups within the same region and at the same time. To absorb unobserved industry-specific trends we add industry times year fixed effects $\theta_i \times \delta_t$, and to address unobserved industry-specific advantages or disadvantages of certain places we add industry times county fixed effects $\theta_i \times \gamma_d$. Since all fixed effects enter the first and second stage of our IV regressions, they should further alleviate concerns with respect to unobserved trends in the attractiveness of certain regions that influence the movement of inventors, e.g. Silicon Valley for computer scientists. The match between industry-specific human capital with industry-specific entrepreneurial activity should also reduce measurement error, so we expect β to be larger when estimated with (4) than with (1).

4.3.4 First-stage instrument plausibility check (using individual -level regressions)

Before applying our instrument at the county or county-industry level to obtain results, we first establish the plausibility of its first stage by investigating the linkage between the historical surname distribution and the geographical mobility of individual inventors. This approach rests on an extensive demographic literature, including the migration of people, social networks and mobility (Rossi, 2013). For example, Piazza et al. (1987) tracks migration rates using surname distribution in Italy, Degioanni & Darlu (2001) infer the geographical origin of migrants in a given area using surnames, and Darlu et al. (2011) show that surname distribution can be used to estimate mobility using the example of Savoy, France. Studies also use surnames to investigate social mobility, e.g., whether social status changes over centuries (Clark & Cummins, 2014) and whether wealth moves over generations (Clark & Cummins, 2015). In a recent study, Grilli & Allesina (2017) perform a surname analysis on academic professors to compare academic systems in the U.S., France, and Italy.

Our IV approach rests on the assumption that historic surname shares can discriminate between destination counties of moving inventors with a given last name, conditional on moving. We empirically test this assumption by estimating a dyadic model that reflects the complete choice set of a moving inventor. To this end, we construct a dataset at the inventor-origin-destination county level that contains each potential destination county combined with the actual county a given inventor is emigrating from. We mark the county the inventor actually moved to with a dummy and for the actual and each potential destination county, include the share of people in the 1940 Census with the same surname. Armed with this dyadic dataset covering 258,657 moves from 1988-2014, we estimate the following model with OLS:

$$Pr(d.cty\#o.cty_{i,o,d,t} = 1|Move\ out_{o,t}) = \alpha_0 + \beta \cdot (\frac{P_{dn}^{1940}}{P_n^{1940}}) + \delta_t + \gamma_d + \gamma_o + \varepsilon_{i,d,o,t} \quad (5)$$

where $Pr(d.cty\#o.cty_{n,o,d,t} = 1|Move\ out_{o,t})$ is a dummy indicating the destination county *d.cty* a given inventor i with name n moved to from origin county *o.cty* in year t. P_{dn}^{1940} is the population in county d with surname n in 1940; P_n^{1940} is the population with surname n in the entire U.S. in 1940; δ_t denotes a full set of year fixed effects to control for varying macroeconomic conditions; γ_d controls for time-invariant unobserved destination county characteristics; and γ_o controls for time-invariant unobserved origin county characteristics that may confound our identification of β, and $\varepsilon_{i,d,o,t}$ is the error term. We estimate four versions of Equation (5): (a) only with year fixed effects; (b) year and destination-county fixed effects; (c) year and origin-county fixed effects; (d) year and destination-origin county combination fixed effects. Variant (d) absorbs time-invariant county-pair relationship characteristics including, for instance, the geographic distance between two counties. Table 4-2 presents the results.

Table 4-2. Destination county choice

	origin-destination county move			
	a	b	c	d
Destination county	0.044***	0.021***	0.044***	0.013***
Historic surname fraction	(0.006)	(0.002)	(0.006)	(0.001)
N	524,583,139	524,583,139	524,583,139	523,553,217
Year FEs	Yes	Yes	Yes	Yes
Destination county FEs	No	Yes	No	No
Origin county FEs	No	No	Yes	No
Origin-destination county FEs	No	No	No	Yes
R^2	0.000	0.008	0.000	0.061

Notes. This table presents OLS regressions of a dummy indicating an origin-destination county move of an inventor within the period 1980-2015 on destination counties' historic surname shares in 1940. Unit of observation is the origin-destination county dyad. Standard errors clustered at the destination county appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

Although we cannot interpret our LPM specification as a probability model, all specifications consistently show that an increase in the historic surname share in a potential destination county leads to a significantly higher probability of observing a given inventor moving to that specific destination county as compared to all other potential destination choices. The results in Table 4-2 support the plausibility of our instrument. The increase in explained variation when destination and destination-origin county fixed effects are included reinforces that unobserved time invariant factors also explain mobility decisions.

4.4 Results

We begin in Table 3 by analyzing the impact of incoming inventors on entrepreneurial quantity. These baseline models regress the logged number of venture-backed startups founded in county d during year t on the logged number of incoming inventors in t-1 (where t-1 is an upper bound of the actual time of arrival, see above for details). Table 4-3, model (a) estimates Equation (1) via naïve OLS. Model (b) applies our IV approach with the instrument defined in (3). Model (c) includes state-year fixed effects to absorb unobserved impacts from US states' policy changes and model (d) adds county fixed effects. Model (e) shows estimates that exclude the top 10 entrepreneurial counties including Silicon Valley.[26]

Interpreting Table 4-3, model (a) shows a strong correlation between the number of incoming inventors in a county with the count of venture-backed startups founded the following year, consistent with Glaeser & Kerr (2009). The remaining models (b-d) employ the IV approach and all show a significant positive impact of incoming inventors in a given county on the local rate of startup formation. The strength of the instrument drops somewhat after the inclusion of county fixed-effects; however, the first stage F value always remains well above conventional levels, suggesting that the IV regression does not suffer from weak instrument bias (Stock & Yogo, 2002; Lee et al., 2021).

---

[26] The top 10 entrepreneurial counties include Alameda County, Los Angeles County, Orange County, San Diego County, San Francisco County, San Mateo County, Santa Clara County in California, Middlesex County in Massachusetts, New York County in New York and King County in Washington.

Table 4-3. Impact of incoming inventors on local venture backed startups

	Venture-backed startups founded				
	a	b	c	d	e
	OLS	IV	IV	IV	IV (w/o top 10 counties)
Incoming Inventors$_{t-1}$	0.360***	0.510***	0.513***	0.180***	0.170***
	(0.019)	(0.027)	(0.027)	(0.040)	(0.040)
N	65,247	65,247	65,247	65,247	65,058
First stage F		804.265	781.375	175.723	139.252
Year FE	Yes	Yes	No	No	No
State FE	Yes	Yes	No	No	No
State-Year FE	No	No	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes
R^2	0.500				

Notes. This table presents OLS regressions of log(number of venture-backed startups + 1). Incoming inventors as well as the instrument are log-transformed. Specifications (b)-(d) show results of our IV regression as described above, where incoming inventors are instrumented with the shift-share instrument (leave-out) in the first stage. Specification (e) show results of our IV regression, but excluding top 10 entrepreneurial counties from the sample. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

In our preferred model (d), the coefficient also drops below the naïve OLS estimate, arguably because the IV reduces bias from self-selection of inventors into more prosperous counties. Model (e) further supports that our results are not limited to Silicon Valley and similar areas. Rather, arriving inventors give rise to more startups generally. Under the assumption that the estimated coefficient can be interpreted as an elasticity, model (d) suggests that a 10% increase in the rate of incoming inventors increases the rate of venture-backed startups founded by 1.8% at the mean.[27] Translating the relative increases into absolute numbers suggests that 10 more inventors lead to 0.035 more startups. Put differently, a county can expect one additional venture-backed startup for every 28.4 incoming inventors.

4.4.1 Technology-specific effect

One concern with the baseline analysis is that the linkage between the arrival of inventors in fields unrelated to the industry where startups are founded, which may add measurement error and downward bias our results. If for example a focal county only had software inventors move in, but all of the increase in startup activity was in biotechnology, we might wonder whether our model accurately enough resembles the notion of an application of task-specific human capital (Gibbons & Waldman, 2004) to relevant new ventures. To this end, we turn to the county-industry level instrument, as described above and formally shown in Equation 4, where inventors are mapped to specific VentureXpert industry categories based on the corresponding technology classes of their patents (Table A2-1 in Appendix 2). The analysis resembles that of Table 4-3, but the dependent variable is the number of startups (models a and b) founded in industry i at a given county c and time t. The finer unit of measurement leads to an increase in the number of observations

[27] It requires a careful interpretation of the estimate. As we did not consider detailed characteristics or types of inventors, the estimate may have been either upward or downward biased. For instance, corporate inventors who are affiliated in a firm may have fewer incentives to found a startup compared to lone inventors if they relish job security or complementary assets within the firm. In contrast, they may also have more chances to spin off and create a startup based on their knowledge developed in the previous firm. Thus, a careful consideration of detailed characteristics is required when interpreting the estimate.

although the underlying data source stays the same. Econometrically it has the advantage of allowing absorption of any unobserved shocks at the county level, whether time variant or invariant, through a richer set of fixed effects: county-industry, county-year, and industry-year.

Table 4-4, model (a), estimates a positive effect on the founding of ventures in the same industry as the inventors of supporting technologies. Table 4-4 model (a) implies that a 10% increase in the rate of incoming inventors increases the rate of venture-backed startup formations in their field by 5.1% at the mean. The larger positive coefficient, relative to Table 4-3 model (d), is consistent with a reduction in measurement error. This result suggests that the findings in Table 4-3 are not spurious due to a generally "rising tide" of startups due to an overall increase in population or supply of technical talent overall; rather, startups arise in the same sectors in which talent has recently been boosted. This supports the inference that an increase in the local supply of technical human capital is causally responsible for entrepreneurial activity in that same sector. It is reminiscent of Bell et al.'s (2020) finding that children are not only more likely to become inventors when they are born in the vicinity of more inventors, but they are more likely to become inventors in the same fields as the inventors they are exposed to.

The field-specific nature of this exposure is further reinforced by model (b), which reveals a negative effect for unrelated technical sectors. This offsetting result makes sense in the context of venture-backed startups, as venture investors must decide how to allocate their dollars. If biotech inventors arrive in the county and biotech startups get funded, it follows that fewer (local) dollars are available for non-biotech startups, as we see in model (b). These results support Lerner & Nanda's (2020) arguments that VCs look for, "…a very narrow band of technological innovations…" (p. 238) and that venture capital reaches a relatively small proportion of entrepreneurial startups.

Table 4-4. Industry-specific inventors and startups

	Venture-backed startups founded	
	a	b
	In same industry	In different industries
	IV	IV
Incoming Inventors$_{t-1}$	0.507***	-0.320***
	(0.052)	(0.033)
N	326,235	326,235
First Stage F	143.955	143.955
County-Industry FE	Yes	Yes
County-Year FE	Yes	Yes
Industry-Year FE	Yes	Yes

Notes. This table presents OLS regressions of log(number of venture-backed startups + 1). All specifications show results of our IV regression as described above, where incoming inventors are instrumented with the shift-share instrument (leave-out) in the first stage. Specifications (a) and (b) present the results for number of venture-backed startups founded in the same and different industries compared to the expertise of incoming inventors, respectively. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

4.4.2 Quality of startups (Successful vs. Failure)

So far, we have established that the arrival of inventors is responsible for the founding of new firms. Although many governments adopt the number of startups as an easy-to-count metric (Lerner, 2009), to truly contribute to jobs, productivity, and growth one would want to measure successful startups. Haltiwanger et al. (2013) note that although startups create many jobs, they also destroy many jobs because failure is the modal outcome. But "success" is not easily discerned. Although Initial Public Offerings almost always indicate a successful startup, acquisitions can be an ambiguous indicator of success. Puri & Zarutski (2012) report that many venture-backed failures are "disguised" as acquisitions, often sold for pennies on the dollar. As noted above, VentureXpert was missing many exit values, so we merged Pitchbook and CrunchBase data with VentureXpert to augment coverage.

In Table 4-5 we only consider the venture-backed startups founded in county d during year t as the dependent variable that become successful within a ten-year window. In model (a), "Successful" is determined retrospectively as the number of firms founded that achieved an IPO or were acquired with a 125% rate of return (as per Ewens & Marx, 2018). The estimates from model (a) suggest that a 10% increase in the rate of incoming inventors increases the rate of successful venture-backed startups founded by 1.0% at the mean. Translating the relative increases into absolute numbers suggests that 10 more inventors lead to 0.022 more successful startups. Put differently, a county can expect one additional successful venture-backed startup for every 460 incoming inventors.

The result in model (a) indicates that incoming inventors are not only responsible for an increase in entrepreneurial activity, as in Table 4-4, but also an upshot in successful startups and assumedly the accompanying jobs, innovations, growth, and liquidity events. One might wonder whether these inventors are only responsible for startups that “just barely” succeeded in returning capital to investors, as opposed to generating some of the more spectacular returns and success stories. We further raised the threshold of an exit value to 500% of total venture capital acquired in model (b), which substantially reduces the magnitude of the estimated coefficient but remains statistically significant. In model (c), we show that inventors even give rise to so-called “unicorn” startups with exit values in excess of 1 billion dollars.

Of course, this increase in the number of successful startups—at all levels—could be a mechanical result of “more shots on goal” so to speak. That is, investors place more bets on more startups and win more often. Therefore, we also test how the influx of inventors affects the failure rate of startups, i.e., venture-backed startups founded in county d during year t that eventually failed. In model (d), we use the traditional measure of “failed” startups as those that are currently Defunct or Bankrupt as indicated in VentureXpert. The results suggest arriving inventors reduce formation of failed startups in the county. Mindful of the Puri & Zarutskie (2012) discovery of failed venture-backed startups “disguised” as acquisitions, in model (e) we include with bankruptcies exits with a valuation lower than 125% of total venture capital invested. Model (e) likewise shows a negative effect of incoming inventors on failed startup foundations (and is robust to eliminating exits with >100% return on investment, or >50%). We conclude that inventors not only causally improve the quantity but also the quality of entrepreneurship.

Table 4-5. Venture-backed startups: Successful vs. Failure

	Successful venture-backed startups			**Failed venture-backed startups**	
	a	b	c	d	e
	Successful (RoR ≥ 125%)	Successful (RoR ≥ 500%)	Successful (Exit ≥ 1B)	Failed	Failed or RoR < 125%
	IV	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.104***	0.068***	0.014**	-0.212***	-0.123***
	(0.033)	(0.023)	(0.006)	(0.028)	(0.027)
N	65,247	65,247	65,247	65,247	65,247
First Stage F	175.723	175.723	175.723	175.723	175.723
State-Year FE	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes

Notes. This table presents OLS regressions of log(number of venture-backed startup foundations + 1). All specifications show results of our IV regression as described above, where incoming inventors are instrumented with the shift-share instrument (leave-out) in the first stage. In specification (a), we define "successful" startups as those that complete either an IPO or successful acquisition within 10 years and achieve a value ≥ 125% of total venture capital acquired. In specification (b), we raised the threshold of an exit value to 500% of total venture capital acquired. In specification (c), we define we define "successful" startups as those that complete either an IPO or successful acquisition within 10 years and achieve an absolute value ≥ 1B dollars, respectively. In specification (d), we define "failed" startups as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. In specification (e), we also include startups that complete either an IPO or successful acquisition within 10 years, but achieve a value < 125% of total venture capital acquired. Incoming inventors as well as the instrument are log-transformed. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

4.4.3 Reallocation from low-tech into high-tech sectors

In Table 4-6 we dig deeper into the dynamics underlying the reallocation in Table 4-5 from lower to higher quality investments. In exploring these mechanisms, we are mindful of past findings that venture investors are local in their investment ability (Sorenson & Stuart, 2001), sensitive even to the availability of direct vs. connecting flights (Bernstein et al., 2016). Therefore, state- and even county-level investment decisions may be influenced by the local supply of inventors. We separate high-tech (biotechnology, life science, computer and communication and semiconductor) from low-tech ventures as defined by VentureXpert. Models (a) and (d), which resemble Table 4-3 in using count of startups as the dependent variable, show a shift from low-tech to high-tech startups upon inventor arrival.

Models (b, c and e, f) of Table 4-6 explore the dynamics of this reallocation from low- to high-tech, breaking down high- and low-tech into Successful vs. Unsuccessful as in models (a) and (d) of Table 5. Model (c) of Table 6 shows a clear shift away from failed low-tech startups. Model (b) shows that successful low-tech startups also decrease in response to arrival of inventors, though the estimated coefficient is much smaller in magnitude than that of failed low-tech startups and also less precisely estimated. This suggests that the shift is primarily away from the failed startups in low-tech industries; in other words, investors appear savvy enough to keep investing in low-tech firms that prove successful, but they avoid less promising low-tech vehicles when inventors arrive.

Models (e) and (f) largely echo the results of Table 4-5, again suggesting that the influx of inventors improves the efficiency of venture investment, reallocating away from failed, low-tech startups toward successful, high-tech startups.

Table 4-6. Venture-backed startups: high-tech vs low-tech, successful vs. unsuccessful

	Low tech			**High tech**		
	a	b	c	d	e	f
	All startups	Successful	Failed	All startups	Successful	Failed
	IV	IV	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	-0.136***	-0.017*	-0.163***	0.356***	0.128***	-0.083***
	(0.029)	(0.010)	(0.022)	(0.042)	(0.034)	(0.018)
N	65,247	65,247	65,247	65,247	65,247	65,247
First Stage F	175.723	175.723	175.723	175.723	175.723	175.723
State-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table presents OLS regressions of log(number of startup foundations + 1) separated by high tech and low tech industries. High- vs. low-tech are categorized according to VentureXpert classifications. Specification (a) and (d) show results of all venture-backed startups foundations. Specification (b) and (e) show results of successful venture-backed startups foundations, where "successful" startups are defined as newly founded venture-backed startups that complete either an IPO or successful acquisition within 10 years and achieve a value ≥ 125% of total venture capital acquired. Specification (c) and (f) show results of failure venture-backed startups foundations, where "failed" startups are defined as those that are currently "Defunct" or "Bankruptcy" as indicated in VentureXpert database. Incoming inventors as well as the instrument are log-transformed. All specifications show results of our IV regression as described above, where incoming inventors are instrumented with the shift-share instrument (leave-out) in the first stage. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

4.4.4 Robustness - Alternative instrument constructions

Although the validity of shift-share instruments does not require exogeneity of the shares, and concerns should be lessened by the inclusion of county fixed effects, we nonetheless estimate robustness checks that should further alleviate concerns of potentially-endogenous share characteristics. We re-estimate model (d) of Table 4-4, replacing the instrument with alternative calculations of the historic name shares (still applying the leave out strategy). Table 4-7 shows the results for these alternative instruments.

For the first alternative instrument (model a), we consider only people in the 1940 Census that lived in a given county before 1935. We thus effectively enlarge the gap between the shares and the actual moves of inventors and reduce potential correlation between historic and current inventor migration shocks. In model b, we exclude the 50 surnames that appear most frequently in the historic data, which should reduce concerns that correlated shares of two counties may lead to an over-rejection problem (as shown by Adao et al., 2019). In our third construction (model c), we exclude wealthy families of each county as inventors may benefit even generations later from their ancestors' wealth. Using the historic house value in the 1940 Census, we excluded families holding more than 1% of the total house value of a given county.

Our fourth construction (model d) departs from the shift-share approach, instead calculating the inventor's separation from their surname's historic geographic centroid. We use the inverse geographic distance between each county centroid and the geographic centroid for an inventor's surname as weights when constructing the instrument. The distance between a county's centroid and a surname's historic geographic centroid has the advantage of a very low correlation with any future county or inventor specific characteristics. A limitation of this fourth instrument construction is that most surnames are clustered in multiple geographic and typically urban regions. Thus, even if there is one largest centroid, we will calculate distance from it even if a somewhat smaller but much-closer aggregation exists. The shift-share instrument does not suffer from this limitation and remains our preferred instrument.

The coefficient sizes remain robust across different specifications, although the strength of the instrument declines in model (d) compared to our original instrument. Especially with respect to our centroid-distance instrument, this is not surprising. That the instrument strength and coefficient size does not decline greatly when excluding particularly influential families supports the assumption that either 1) there is no direct link between the historic name shares and the second stage regression, or 2) the county fixed effects effectively absorb such potentially worrying relationships.

4.4.1 Placebo tests: random reassignment of instrument

Given the relative strength of the instrument, one might wonder whether our IV effectively absorbs unobserved local characteristics and hence leads to an overly strong rejection of the null hypothesis. To address these concerns, we run three placebo tests in the spirit of Adao, Kolesár, & Morales (2019, henceforth AKM). We randomly reassign

Table 4-7. Alternative instruments

	Successful venture-backed startups founded			
	a	b	c	d
	Only individuals settled by 1935	Dropped 50 most frequent surnames	Dropped wealthy families	Alternative instrument using centroid
	IV	IV	IV	IV
Incoming Inventors$_{t-1}$	0.110***	0.109***	0.106***	0.272**
	(0.036)	(0.035)	(0.033)	(0.109)
N	65,247	65,247	65,247	65,247
First Stage F	159.068	162.303	183.076	22.971
State-Year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes

Notes. This table presents OLS regression of log(number of successful venture-backed startups founded + 1), where "successful" startups are defined as newly found venture backed companies that complete either an IPO or successful acquisition within 10 years and achieve a value > 125% of total venture capital acquired. Incoming inventors as well as the instrument are log-transformed. Model (a) restricts the instrument to those who settled in the county of the 1940 Census by 1935; (b) excludes the 50 most frequent surnames; (c) excludes the wealthiest 1% of surnames per 1940 Census house value; (d) replaces the shift-share approach with the inverse geographic distance between the county and the centroid for the inventor's surname. First stage F is the Kleibergen-Paap Wald F statistic of the first stage regression. Standard errors clustered at the county level appear in parentheses. ***, ** and * indicate a significance level of 1%, 5%, and 10%, respectively.

the instrument in three ways: (1) across the entire sample, (2) across counties within a given year, and (3) across time within a given county. Then, we re-run our baseline model with each placebo 1000 times. Table 4-8 summarizes the results of the first and second stages. All three placebos consistently show that a random assignment effectively eliminates a significant prediction of incoming inventors in the first stage, and false identification of a causal impact of incoming inventors on the number of successful venture-backed startups in the second stage. Hence, our IV estimates do not seem to suffer from the artificial over-rejection of the null hypothesis as identified in many other applications of shift-share instruments by AKM. The reason would seem to lie in the effective absorption of unobserved time-invariant heterogeneity at the county level.

Table 4-8. Results from placebo analysis

	a	b	c	d
	Coefficient		Std. Err.	Rejection rate
	(Mean)	(Std. Dev.)	(Median)	(%)
Panel A: Placebo IV randomly shuffled across the overall sample				
1st stage	0.000	0.002	0.002	5.5
2nd stage	2.223	67.497	0.598	0.0
Panel B: Placebo IV randomly shuffled across counties within each year				
1st stage	0.000	0.002	0.002	5.2
2nd stage	0.090	11.871	0.638	0.1
Panel C: Placebo IV randomly shuffled across years within each county				
1st stage	-0.004	0.008	0.007	8.5
2nd stage	-0.264	4.164	0.742	0.1

Notes. We randomly shuffle our instrument to construct placebo instrument variables across the overall sample (Panel A), across counties within each year (Panel B), and across years within each county (Panel C). For each placebo instrument variables, we ran 1000 regressions of log(number of successful venture-backed startup foundation + 1) on incoming inventors, instrumented with the placebo IV that is newly generated for each regression. Incoming inventors as well as the placebo instrument are log-transformed. Column (a) and (b) report the mean and standard deviation of the coefficients obtained from 1000 placebo regressions, respectively. Column (c) reports the median value of the standard error for the coefficient of each regression over 1000 placebo regressions. Column (d) reports the rate of which the regression rejects the null hypothesis of no effect at the 5% significance level over 1000 placebo regressions. We report these values corresponding to each of the first and second stages of the placebo regressions.

4.5 Conclusion

We have provided arguably causal evidence regarding how the arrival of inventors influences both the quantity and quality of entrepreneurship. Our shift-share instrument, based on the county-level distribution of surnames in the 1940 U.S. Census, addresses limitations of similar instruments in the international-migration literature. We are able to show a sector-specific uptick in entrepreneurial activity and also tie the arrival of inventors to a rise in successful startups as well as a lowering of unsuccessful startups. Our estimates indicate that approximately 460 new inventors in a county can create a successful startup, and even "unicorn" startups with >$1B exits can be traced to inventor arrivals. The approach further illustrated how venture capital firms shifted their investment towards high technology opportunities, at the expense of unsuccessful low technology opportunities. The shift away from unsuccessful low tech to high tech firm starts held across all U.S. counties—not just Silicon Valley and similar hotspots—as well as a variety of instruments, and measures.

Although this work sought to explain how the supply of inventors influenced high-growth entrepreneurship, it can also speak to the classic question of why industries cluster geographically (Rosenthal & Strange, 2004; Overman & Puga, 2010; Ellison et al., 2010). Much work has validated the Marshallian agglomeration arguments of production economies, labor pooling, and knowledge spillovers, yet that work has often struggled to isolate and estimate causal mechanisms (Glaeser & Kerr 2009). The shift share instrument developed here enabled investigation of one arguably causal linkage; inventor arrival fuels an increase and funding in startups in those inventors' specific industries. Furthermore, if inventors move towards incipient clusters (e.g., semiconductors in Silicon Valley in the 1960s), their impact on field-specific entrepreneurship and venture capital investment could create a feedback dynamic that directly and dramatically fuels industry concentration.

Given the increasing importance of technology, innovation, and the growth of the knowledge economy, these results also imply an ever-increasing role for STEM labor pooling amongst the three classic Marshallian mechanisms. Assuming that inventor immigration to a region bolsters this role, these results would imply that pooling drives investment which could in turn result in the co-location of production assets. Given that knowledge spillovers are localized and probably reliant upon personal inventor communication (Saxenian, 1996; Thompson and Fox-Kean 2005), then inventor pooling should also increase knowledge spillovers. Future research should seek to disentangle the Marshallian mechanisms that drive agglomeration, estimate their feedback effects, and quantify their relative importance.

The mutual reinforcement of these agglomeration mechanisms could partially explain the rapid emergence of Silicon Valley and the growth in inequality across regions in the U.S. (Glaeser & Hausman, 2020; Lerner & Nanda, 2020). Moretti (2012) labels this phenomenon the "Great Divergence" and provides an example of two relatively similar California towns in 1969 – Menlo Park and Visalia. Surprisingly, given their wide differences now across wealth, crime, education, and health measures, the towns had relatively similar incomes and educational levels in 1969. The venture capital firm Kleiner-Perkins founded their operations in Menlo Park in 1972 and became prominent after a series of high-profile successes, including Amazon, Google, and Genentech. Their private success and similar successes by other nearby investors created a striking concentration of wealth (Lerner & Nanda, 2020), for example, for many years, real estate on Menlo Park's Sand Hill road was the most expensive in the world.

Independent from its implications for regional inequality, this work enables a crude estimate of the "value" of an inventor; geography and mobility in this respect simply provide an instrument to get at that estimate. This estimate is obviously sensitive to the region in which it is derived; the value of an inventor surely varies across regions, based on the inventor, the region, and the interaction of the two. Although this work used arrival in a county to back out the value of an arriving inventor, a home-grown inventor might be just as useful to local entrepreneurship (for example, Steve Wozniak already lived and worked in Silicon Valley before founding Apple). Indeed, if a home-grown inventor had easier access to existing networks of friends, family, investors, and fellow entrepreneurs, they might be even more effective at supporting the success of high-tech firms. It would be interesting to explore whether inventor arrival crowds out–or complements—locally grown inventors and entrepreneurship (Azoulay et. al. 2021).

Though beyond the scope of this work, this study leaves several possibilities for future research. This study does not consider specific characteristics of inventors and estimate the heterogeneous effect of inventors with different characteristics. Inventors with different characteristics, for instance, by type of their expertise (e.g., Arts & Fleming, 2018), individual characteristics (e.g., Zwick et al., 2017), affiliation status (e.g., Singh & Fleming, 2010), or intangible assets (e.g., Paruchuri & Eisenman, 2012), would have a differential impact on regional entrepreneurship. These are not included in this study as the purpose of this work is to obtain an accurate estimate of the value of inventors in terms of entrepreneurship using the novel shift-share instrument. Thus, future work could explore the heterogeneous effect of inventors depending on their individual characteristics. The causal effect of startup activities in a region on inventor mobility to/from the region is another important research gap to address. Though this study rules out the issue of reverse causality in estimating the causal impact of inventors on entrepreneurship, the foundation of startups and the career opportunities would influence the mobility of inventors. Cheyre et al. (2015) showed how spinoffs promote mobility of inventors moving from incumbents to recent entrants focusing on the semiconductor industry in Silicon Valley. Since regional entrepreneurship is not limited to spinoffs, a further investigation is required to fully understand the influence of startup activities.

Future work could also estimate how the loss of inventors impacts the source region. Our back of the envelope calculations implied that inventor arrivals enable 17.9% of high-tech entrepreneurship, however, this calculation ignores the probably negative impact on the home regions of the arrivals. Although beyond the scope of this work, a full accounting of these effects might enable an estimate of the social welfare of inventor mobility. This could then inform policy, for example, should policies encourage industries and technologies to cluster, because such clustering improves innovative efficiency (for one example, through increases in knowledge spillovers), or should policies encourage industries to disperse, and hence distribute jobs and wealth in a more geographically equitable way?

Beyond inventor pooling and investment, regional entrepreneurship ecosystems also depend on physical and institutional infrastructure, lawyers, and non-technical entrepreneurial talent. There are surely declining marginal returns as the supply of tech talent outstrips complementary resources needed for entrepreneurship. This can be seen in the negative effects inventors in one field have on the financing of startups in other fields. Future work should investigate whether the arrival of an inventor in one county decreases entrepreneurship in nearby counties, possibly due to competition for complementary resources.

Chapter 5. Conclusion

The dissertation extends prior literature on innovation by demonstrating how scientific knowledge and human capital are key components of innovation and economic growth.

Using a data sample comprising scientific discovery in US FFRDCs, Chapter two of this dissertation shows how the patents filed by government scientists influence the dissemination of scientific discovery and follow-on inventions based on scientific knowledge. It finds a means of which help disseminate scientific discoveries in government laboratories, i.e., patent filing by the responsible government scientists. This finding provides important practical implications for policymakers regarding government laboratories, especially those aiming to create a complementary structure between government and industrial laboratories.

Chapter three of this dissertation explores the conditions that influence mobility decision of inventors. This chapter demonstrates that historic surname distribution influences the geographical mobility of inventors and how it interacts with other families, inventors, and institutional factors to shape the mobility decision. It provides practical guidance for policymakers to attract or retain inventors in their locations.

Chapter four of this dissertation presents the significant role of inflowing inventors in regional entrepreneurship. The newly developed shift-share instrument allows the estimation of the effect of incoming inventors on startup activities at the county level.

Empirical evidence on various aspects of surname effects in Chapter three supports the plausibility of the share part of the shift-share instrument. Strengthening the causality with the instrument, Chapter four estimates the effect size of inventors on the rate and quality of startup foundations based on comprehensive data on inventors and venture-backed startups in the US. This provides empirical evidence to understand how Silicon Valley was possible, and thus help policymakers plan for a regional innovation ecosystem.

The dissertation leaves directions for future studies in the field of innovation. Chapter two of the dissertation examines only one of the many means that may help disseminate scientific discovery in government laboratories. Considering the importance of dissemination and utilization of government science, there are still many research possibilities around the means to diffuse government science, e.g., collaboration of scientists, conference participation, etc. Chapter three and four of the dissertation develop and test the plausibility of the shift-share instrument for the geographical mobility of inventors. This instrument can aid future studies that estimate the causal impact of inventors or other types of human capital on regional outcomes, opening up possibilities for estimating the value of human capital in various aspects.

Bibliography

Adams, J. D., Chiang, E. P., & Jensen, J. L. (2003). The influence of federal laboratory R&D on industrial research. Review of Economics and Statistics, 85(4), 1003-1020.

Adao, R., Kolesár, M., & Morales, E. (2019). Shift-share designs: Theory and inference. The Quarterly Journal of Economics, 134(4), 1949-2010.

Agarwal, R., Ganco, M., & Ziedonis, R. H. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. Strategic Management Journal, 30(13), 1349-1374.

Agrawal, A., Goldfarb, A., & Teodoridis, F. (2016). Understanding the changing structure of scientific inquiry. American Economic Journal: Applied Economics, 8(1), 100-128.

Agrawal, A., Kapur, D., McHale, J., & Oettl, A. (2011). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. Journal of Urban Economics, 69(1), 43-55.

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. Science, 357(6351), 583-587.

Akcigit, U., Baslandze, S., & Stantcheva, S. (2016). Taxation and the international mobility of inventors. American Economic Review, 106(10), 2930-2981.

Aldridge, T. T., & Audretsch, D. (2011). The Bayh-Dole Act and scientist entrepreneurship. Research Policy, 8(40), 1058-1067.

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. Management science, 45(7), 905-917.

Arora, A., Belenzon, S., & Patacconi, A. (2018). The decline of science in corporate R&D. Strategic Management Journal, 39(1), 3-32.

Arora, A., Belenzon, S., & Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. American Economic Review, 111(3), 871-98.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention (pp. 609-626). Universities-National Bureau Committee for Economic and Council (ed.), The Rate and Direction of Inventive Activity: Economic and Social Factors, Princeton.

Arrow, K. J., & Capron, W. M. (1959). Dynamic shortages and price rises: the engineer-scientist case. The Quarterly Journal of Economics, 73(2), 292-308.

Arts, S., & Fleming, L. (2018). Paradise of novelty—or loss of human capital? Exploring new fields and inventive output. Organization Science, 29(6), 1074-1092.

Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does science advance one funeral at a time? American Economic Review, 109(8), 2889-2920.

Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2020). Age and high-growth entrepreneurship. American Economic Review: Insights, 2(1), 65-82.

Azoulay, P., Jones, B., Kim, J. D., & Miranda, J. (2021). Immigration and Entrepreneurship in the United States. American Economic Review, Forthcoming.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., ... & Fleming, L. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. Journal of Economics & Management Strategy, 27(3), 535-553.

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G. C., Lück, S., O'Reagan, D., & Yeh, B. (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. Journal of Economics & Management Strategy, 27(3), 535-553.

Balsmeier, B., Fleming, L., & Manso, G. (2017). Independent boards and innovation. Journal of Financial Economics, 123(3), 536-557.

Bartel, A. P. (1989). Where do the new US immigrants live?. Journal of Labor Economics, 7(4), 371-391.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?.

Baruffaldi, S. H., & Simeth, M. (2020). Patents and knowledge diffusion: The effect of early disclosure. Research Policy, 49(4), 103927.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. Strategic organization, 6(3), 285-327.

Baumol, William J. and Bowen, William G. (1966) Performing Arts: The Economic Dilemma. MIT Press, Cambridge.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who becomes an inventor in America? The importance of exposure to innovation. The Quarterly Journal of Economics, 134(2), 647-713.

Bernstein, S., Diamond, R., McQuade, T., & Pousada, B. (2018). The contribution of high-skilled immigrants to innovation in the United States. Stanford Graduate School of Business Working Paper, 3748, 202019-20.

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. The Journal of Finance, 71(4), 1591-1622.

Bhaskarabhatla, A., Cabral, L., Hegde, D., & Peeters, T. (2021). Are inventors or firms the engines of innovation?. Management Science, 67(6), 3899-3920.

Bikard, M., & Marx, M. (2020). Bridging academia and industry: How geographic hubs connect university science and corporate technology. Management Science, 66(8), 3425-3443.

Borusyak, K., Hull, P., & Jaravel, X. (2018). Quasi-experimental shift-share research designs (No. w24997). National Bureau of Economic Research.

Bozeman, B. (2000). Technology transfer and public policy: a review of research and theory. Research Policy, 29(4-5), 627-655.

Bozeman, B., & Crow, M. (1991). Technology transfer from US government and university R&D laboratories. Technovation, 11(4), 231-246.

Bozeman, B., Rimes, H., & Youtie, J. (2015). The evolving state-of-the-art in technology transfer research: Revisiting the contingent effectiveness model. Research Policy, 44(1), 34-49.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. Journal of economic geography, 9(4), 439-468.

Breschi, S., Lissoni, F., & Miguelez, E. (2017). Foreign-origin inventors in the USA: testing for diaspora and brain gain effects. Journal of Economic Geography, 17(5), 1009-1038.

Burchardi, K. B., Chaney, T., Hassan, T. A., Tarquinio, L., & Terry, S. J. (2020). Immigration, innovation, and growth (No. w27075). National Bureau of Economic Research.

Cappelli, R., Czarnitzki, D., Doherr, T., & Montobbio, F. (2019). Inventor mobility and productivity in Italian regions. Regional Studies, 53(1), 43-54.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. Journal of Labor Economics, 19(1), 22-64.

Chemmanur, T. J., Kong, L., Krishnan, K., & Yu, Q. (2019). Top management human capital, inventor mobility, and corporate innovation. Journal of Financial and Quantitative Analysis, 54(6), 2383-2422.

Cheyre, C., Klepper, S., & Veloso, F. (2015). Spinoffs and the mobility of US merchant semiconductor inventors. Management Science, 61(3), 487-506.

Clark, G., & Cummins, N. (2014). Surnames and social mobility in England, 1170–2012. Human Nature, 25(4), 517-537.

Clark, G., & Cummins, N. (2015). Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility. The Economic Journal, 125(582), 61-85.

Cockburn, I. M., Henderson, R., & Stern, S. (1999). The diffusion of science-driven drug discovery: Organizational change in pharmaceutical research (No. w7359).

National Bureau of Economic Research.

Conti, R., Gambardella, A., & Mariani, M. (2014). Learning to be Edison: Inventors, organizations, and breakthrough inventions. Organization Science, 25(3), 833-849.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: A bad mix. Technical Note, Duke University.

Crow, M., & Bozeman, B. (1998). Limited by design: R&D laboratories in the US national innovation system: Columbia University Press.

Czarnitzki, D., & Toole, A. A. (2011). Patent protection, market uncertainty, and R&D investment. The Review of Economics and Statistics, 93(1), 147-159.

Darlu, P., Brunet, G., & Barbero, D. (2011). Spatial and temporal analyses of surname distributions to estimate mobility and changes in historical demography: the example of Savoy (France) from the eighteenth to the twentieth century. In Navigating time and space in population studies (pp. 99-113). Springer, Dordrecht.

Darlu, P., Brunet, G., & Barbero, D. (2011). Spatial and temporal analyses of surname distributions to estimate mobility and changes in historical demography: the example of Savoy (France) from the eighteenth to the twentieth century. In Navigating time and space in population studies (pp. 99-113): Springer.

Decker, R., Haltiwanger, J., Jarmin, R., & Miranda, J. (2014). The role of entrepreneurship in US job creation and economic dynamism. Journal of Economic Perspectives, 28(3), 3-24.

Degioanni, A., & Darlu, P. (2001). A Bayesian approach to infer geographical origins of migrants through surnames. Annals of Human biology, 28(5), 537-545.

Dorn, D., Hanson, G. H., Pisano, G., & Shu, P. (2020). Foreign competition and domestic innovation: Evidence from US patents. American Economic Review: Insights, 2(3), 357-74.

El Ghaoui, L., Pham, V., Li, G. C., Duong, V. A., Srivastava, A., & Bhaduri, K. (2013). Understanding large text corpora via sparse machine learning. Statistical Analysis and Data Mining: The ASA Data Science Journal, 6(3), 221-242.

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. American Economic Review, 100(3), 1195-1213.

Ewens, M., & Marx, M. (2018). Founder replacement and startup performance. The Review of Financial Studies, 31(4), 1532-1565.

Fagerberg, J., & Srholec, M. (2008). National innovation systems, capabilities and economic development. Research Policy, 37(9), 1417-1435.

Federal Laboratory Consortium for Technology Transfer. (2019). Federal Technology Transfer Legislation and Policy The Green Book (6th ed.).

Federal Technology Transfer Act of 1986, Pub. L. No.99-502, 15 U.S.C. § 3701, 100 Stat. 1785 (1986).

Fini, R., Rasmussen, E., Siegel, D., & Wiklund, J. (2018). Rethinking the commercialization of public science: From entrepreneurial outcomes to societal impacts. Academy of Management Perspectives, 32(1), 4-20.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. Strategic Management Journal, 25(8-9), 909-928.

Fleming, L., Greene, H., Li, G., Marx, M., & Yao, D. (2019). Government-funded research increasingly fuels innovation. Science, 364(6446), 1139-1141.

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. Administrative science quarterly, 52(3), 443-475.

Galasso, A., & Schankerman, M. (2015). Patents and cumulative innovation: Causal evidence from the courts. The Quarterly Journal of Economics, 130(1), 317-369.

Ganco, M. (2013). Cutting the Gordian knot: The effect of knowledge complexity on employee mobility and entrepreneurship. Strategic Management Journal, 34(6), 666-686.

Gans J., Hsu D., & Stern S. (2008) The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. Management Science, 54(5), 982–997.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human capital and regional development. The Quarterly Journal of Economics, 128(1), 105-164.

Gibbons, R., & Waldman, M. (2004). Task-specific human capital. American Economic Review, 94(2), 203-207.

Glaeser, E. L., & Hausman, N. (2020). The spatial mismatch between innovation and joblessness. Innovation Policy and the Economy, 20(1), 233-299.

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?. Journal of Economics & Management Strategy, 18(3), 623-663.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. Review of Economics and Statistics, 97(2), 498-520.

Goldin, C., & Shim, M. (2004). Making a name: Women's surnames at marriage and beyond. Journal of Economic Perspectives, 18(2), 143-160.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586-2624.

Graham, S., & Hegde, D. (2015). Disclosing patents' secrets. Science, 347(6219), 236-237.

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. Proceedings of the National Academy of Sciences, 114(29), 7600-7605.

Gruber, M., Harhoff, D., & Hoisl, K. (2013). Knowledge recombination across technological boundaries: Scientists vs. engineers. Management Science, 59(4), 837-851.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498). National Bureau of Economic Research.

Hall, R. E., & Woodward, S. E. (2010). The burden of the nondiversifiable risk of entrepreneurship. American Economic Review, 100(3), 1163-94.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. Review of Economics and Statistics, 95(2), 347-361.

Hathaway, I. (2018). High-growth firms and cities in the US: an Analysis of the Inc. 5000. Brookings Institution report.

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship (No. t0017). National Bureau of Economic Research.

Hegde, D., & Luo, H. (2018). Patent publication and the market for ideas. Management Science, 64(2), 652-672.

Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. Science, 280(5364), 698-701.

Hoisl, K. (2007). Tracing mobile inventors—the causality between inventor mobility and inventor productivity. In A Study of Inventors (pp. 65-119): Springer.

Hoisl, K. (2009). Does mobility increase the productivity of inventors? The Journal of Technology Transfer, 34(2), 212-225.

Hombert, J., & Matray, A. (2017). The real effects of lending relationships on innovative firms and inventor mobility. The Review of Financial Studies, 30(7), 2413-2445.

Hunt, J. (2017). The impact of immigration on the educational attainment of natives. Journal of Human Resources, 52(4), 1060-1118.

Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation?.

American Economic Journal: Macroeconomics, 2(2), 31-56.

Iacus, S. M., King, G., & Porro, G. (2009). CEM: software for coarsened exact matching.

Jaeger, D. A., Ruist, J., & Stuhler, J. (2018). Shift-share instruments and the impact of immigration (No. w24285). National Bureau of Economic Research.

Jaffe, A. B., & Lerner, J. (2001). Reinventing public R&D: Patent policy and the commercialization of national laboratory technologies. RAND Journal of Economics, 167-198.

Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. American Economic Review, 90(2), 215-218.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. The Quarterly Journal of Economics, 108(3), 577-598.

Jensen, R., & Thursby, M. (2001). Proofs and prototypes for sale: The licensing of university inventions. American Economic Review, 91(1), 240-259.

Jung, H. J., & Lee, J. J. (2016). The quest for originality: A new typology of knowledge search and breakthrough inventions. Academy of Management Journal, 59(5), 1725-1753.

Kenney, M., & Patton, D. (2009). Reconsidering the Bayh-Dole Act and the current university invention ownership model. Research Policy, 38(9), 1407-1422.

Kerr, W. R. (2013). US high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence (No. w19377). National Bureau of Economic Research.

Kerr, W. R., & Lincoln, W. F. (2010). The supply side of innovation: H-1B visa reforms and US ethnic invention. Journal of Labor Economics, 28(3), 473-508.

Kline, P., Petkova, N., Williams, H., & Zidar, O. (2019). Who profits from patents? rent-sharing at innovative firms. The Quarterly Journal of Economics, 134(3), 1343-

1404.

Kline, S. J., & Rosenberg, N. (1986). An overview of innovation. The positive sum strategy: Harnessing technology for economic growth. The National Academy of Science, USA, 35, 36.

Kortum, S., & Lerner, J. (2001). Does venture capital spur innovation?. Emerald Group Publishing.

LaLonde, R.J. and Topel, R.H. (1991). Labor market adjustments to increased immigration. In Immigration, trade, and the labor market (pp. 167-199). University of Chicago Press.

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. R. (2021). Valid t-ratio Inference for IV (No. w29124). National Bureau of Economic Research.

Lee, K. (2013). Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap. Cambridge University Press.

Lerner, J. (2002). 150 years of patent protection. American Economic Review, 92(2), 221-225.

Lerner, J. (2009). Boulevard of broken dreams: why public efforts to boost entrepreneurship and venture capital have failed and what to do about it. Princeton University Press.

Lerner, J., & Nanda, R. (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. Journal of Economic Perspectives, 34(3), 237-61.

Li, D., Azoulay, P., & Sampat, B. N. (2017). The applied value of public investments in biomedical research. Science, 356(6333), 78-81.

Li, G.-C., Lai, R., D’Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., Amy, Z. Y., & Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). Research Policy, 43(6), 941-955.

Link, A. N., Siegel, D. S., & Van Fleet, D. D. (2011). Public science and public innovation: Assessing the relationship between patenting at US National Laboratories and the Bayh-Dole Act. Research Policy, 40(8), 1094-1099.

Lück, S., Balsmeier, B., Seliger, F., & Fleming, L. (2020). Early disclosure of invention and reduced duplication: An empirical test. Management Science, 66(6), 2677-2685.

Maloney, W. F., & Caicedo, F. V. (2016). The persistence of (subnational) fortune. The Economic Journal, 126(598), 2363-2401.

Marx, M., & Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal, 41(9), 1572-1594.

Marx, M., & Hsu, D. H. (2021). Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries. Management Science, Forthcoming.

Marx, M., Singh, J., & Fleming, L. (2015). Regional disadvantage? Employee non-compete agreements and brain drain. Research Policy, 44(2), 394-404.

Marx, M., Strumsky, D., & Fleming, L. (2009). Mobility, skills, and the Michigan non-compete experiment. Management science, 55(6), 875-889.

Melero, E., Palomeras, N., & Wehrheim, D. (2020). The effect of patent protection on inventor mobility. Management Science, 66(12), 5485-5504.

Moreira, S., & Soares, T. J. (2020). Academic spill-ins or spill-outs? Examining knowledge spillovers of university patents. Industrial and Corporate Change, 29(5), 1145-1165.

Moretti, E. (2012). The new geography of jobs. Houghton Mifflin Harcourt.

Moretti, E., & Wilson, D. J. (2014). State incentives for innovation, star scientists and jobs: Evidence from biotech. Journal of Urban Economics, 79, 20-38.

Moretti, E., & Wilson, D. J. (2017). The effect of state taxes on the geographical location of top earners: Evidence from star scientists. American Economic Review, 107(7), 1858-1903.

Moser, P., Voena, A., & Waldinger, F. (2014). German Jewish émigrés and US invention. American Economic Review, 104(10), 3222-55.

Mowery, D. C. (1992). The US national innovation system: origins and prospects for change. Research Policy, 21(2), 125-144.

Mowery, D. C., Nelson, R. R., Sampat, B. N., & Ziedonis, A. A. (2001). The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole act of 1980. Research policy, 30(1), 99-119.

Mowery, D. C., & Ziedonis, A. A. (2015). Markets versus spillovers in outflows of university research. Research Policy, 44(1), 50-66.

Myers, K. (2020). The elasticity of science. American Economic Journal: Applied Economics, 12(4), 103-34.

NSB, N. (2020). Research and Development: US Trends and International Comparisons. Science and Engineering Indicators 2020.

National Academies of Sciences, Engineering, and Medicine. (2021). Advancing Commercialization of Digital Products from Federal Laboratories. Washington, DC: The National Academies Press.

Nelson, R. R. (1959). The simple economics of basic scientific research. Journal of political economy, 67(3), 297-306.

Nelson, R. R. (1993). National innovation systems: a comparative analysis: Oxford University Press.

Oettl, A., & Agrawal, A. (2008). International labor mobility and knowledge flow externalities. Journal of international business studies, 39(8), 1242-1260.

Overman, H. G., & Puga, D. (2010). Labor pooling as a source of agglomeration: An empirical investigation. In Agglomeration economics (pp. 133-150), edited by Edward Glaeser. University of Chicago Press.

Owen-Smith, J. (2001). Managing laboratory work through skepticism: Processes of evaluation and control. American Sociological Review, 427-452.

Palangkaraya, A., Webster, E., & Jensen, P. H. (2011). Misclassification between patent offices: evidence from a matched sample of patent applications. Review of Economics and Statistics, 93(3), 1063-1075.

Palomeras, N., & Melero, E. (2010). Markets for inventors: learning-by-hiring as a driver of mobility. Management Science, 56(5), 881-895.

Parey, M., Ruhose, J., Waldinger, F., & Netz, N. (2017). The selection of high-skilled emigrants. Review of Economics and Statistics, 99(5), 776-792.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in US cities. Journal of Labor Economics, 33(S1), S225-S255.

Paruchuri, S., & Eisenman, M. (2012). Microfoundations of firm R&D capabilities: A study of inventor networks in a merger. Journal of Management Studies, 49(8), 1509-1535.

Piazza, A., Rendine, S., Zei, G., Moroni, A., & Cavalli-Sforza, L. L. (1987). Migration rates of human populations from surname distributions. Nature, 329(6141), 714-716.

Poege, F., Harhoff, D., Gaessler, F., & Baruffaldi, S. (2019). Science quality and the value of inventions. Science Advances, 5(12), eaay7323.

Polidoro Jr, F., & Yang, W. (2021). Corporate Investment Relationships and the Search for Innovations: An Examination of Startups' Search Shift Toward Incumbents. Organization Science, 32 (4), 909-939.

Puri, M., & Zarutskie, R. (2012). On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms. The Journal of Finance, 67(6), 2247-2293.

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. Management Science, 59(2), 504-525.

Rosenberg, N., & Nelson, R. R. (1994). American universities and technical advance in industry. Research policy, 23(3), 323-348.

Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. Management science, 49(6), 751-766.

Rosenthal, S.S. and Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In Handbook of regional and urban economics (Vol. 4, pp. 2119-2171). Elsevier.

Rossi, A. (2013). Does economic upgrading lead to social upgrading in global production networks? Evidence from Morocco. World Development, 46, 223-233.

Rossi, P. (2013). Surname distribution in population genetics and in statistical physics. Physics of life reviews, 10(4), 395-415.

Saavedra, P., & Bozeman, B. (2004). The "Gradient Effect" in Federal Laboratory-Industry Technology Transfer Partnerships. Policy Studies Journal, 32(2), 235-252.

Saxenian, A. (1996) Regional advantage. Harvard University Press.

Schumpeter, J. (1942) Capitalism, socialism and democracy. Vol. 36, Harper & Row, New York.

Siegel, D. S., Veugelers, R., & Wright, M. (2007). Technology transfer offices and commercialization of university intellectual property: performance and policy implications. Oxford Review of Economic Policy, 23(4), 640-660.

Singh, J., & Agrawal, A. (2011). Recruiting for ideas: How firms exploit the prior inventions of new hires. Management Science, 57(1), 129-150.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. Management science, 56(1), 41-56.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J., & Wang, K. (2015). An overview of microsoft academic service (mas) and applications. Paper presented at the Proceedings of the 24th international conference on world wide web.

Smith, A. (1776). Wealth of nations.

Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. Research

Policy, 33(10), 1615-1634.

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. American Journal of Sociology, 106(6), 1546-1588.

Sorenson, O., Assenova, V., Li, G.-C., Boada, J., & Fleming, L. (2016). Expand innovation finance via crowdfunding. Science, 354(6319), 1526-1528.

Starr, E., Balasubramanian, N., & Sakakibara, M. (2018). Screening spinouts? How noncompete enforceability affects the creation, growth, and survival of new firms. Management Science, 64(2), 552-572.

Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression. (No. w0284). National Bureau of Economic Research.

Stuart, T. E., & Sorenson, O. (2003). Liquidity events and the geographic distribution of entrepreneurial activity. Administrative Science Quarterly, 48(2), 175-201.

Thompson, P., & Fox-Kean, M. (2005). Patent citations and the geography of knowledge spillovers: A reassessment. American Economic Review, 95(1), 450-460.

Thursby, J. G., & Thursby, M. C. (2003). University licensing and the Bayh-Dole Act. Science, 301(5636), 1052.

U.S. Department of Energy. (2020). The State of the DOE National Laboratories (2020 ed.). Washington, DC.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. Science, 342(6157), 468-472.

Wagner, S., & Goossen, M. C. (2018). Knowing me, knowing you: Inventor mobility and the formation of technology-oriented alliances. Academy of Management Journal, 61(6), 2026-2052.

Woo, S., Jang, P., & Kim, Y. (2015). Effects of intellectual property rights and patented knowledge in innovation and industry value added: A multinational empirical analysis of different industries. Technovation, 43, 49-63.

Wozniak, A., & Murray, T. J. (2012). Timing is everything: Short-run population impacts of immigration in US cities. Journal of Urban Economics, 72(1), 60-78.

Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. (2021). Science as a Public Good: Public Use and Funding of Science (No. w28748). National Bureau of Economic Research.

Younge, K. A., Tong, T. W., & Fleming, L. (2015). How anticipated employee mobility affects acquisition likelihood: Evidence from a natural experiment. Strategic Management Journal, 36(5), 686-708.

Ziedonis, R. H. (2004). Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. Management Science, 50(6), 804-820.

Zucker, L., Darby, M., & Brewer, M. (1998). Intellectual human capital and the birth of US biotechnology enterprises. American Economic Review, 88(1), 290-306.

Zwick, T., Frosch, K., Hoisl, K., & Harhoff, D. (2017). The power of individual-level drivers of inventive performance. Research policy, 46(1), 121-137.

Appendix 1: Appendix for Chapter two

Rule-based text matching algorithm to classify the origin of science papers

We use data on scientific papers provided by the Microsoft Academic Graph (MAG) to identify scientific papers that originate in each type of laboratory: 1) US FFRDCs, 2) government laboratories in general, and 3) industrial laboratories. The MAG database provides information on the authors and the corresponding affiliations for each paper-author. The affiliation information is provided in a raw string format, which includes the names and addresses of affiliated institutions. Thus, to classify the origin types of scientific papers, it is necessary to develop a rule-based text matching algorithm that captures the institution information from the raw string and identifies each scientific paper as originating from a government laboratory.

1) Identifying scientific papers of US FFRDCs

We start by making a dictionary of possible names for each FFRDCs based on the master list of US FFRDCs, which has been maintained by the National Science Foundation of the United States. We include the official name of FFRDC, previous name, and alternative names, such as abbreviations. Table A1 provides the list of FFRDCs and their alternative names. We match the names on the entire string with a strict restriction applied for the string before and after the word. We also consider possible alternatives for the word ‘Laboratory’ (i.e., ‘Lab’, ‘Labs’, ‘Laboratories’) and ‘Observatory’ (i.e., ‘Observatories’) in the name of FFRDCs.

Table A1-1. Name of FFRDCs

Num.	Official name	Previous names	Alternative Names
0	Aerospace Federally Funded Research and Development Center		Aerospace FFRDC
1	Ames Laboratory		
2	Argonne National Laboratory		

3	Arroyo Center		
4	Brookhaven National Laboratory		
5	National Security Engineering Center	C3I Federally Funded Research and Development Center	C3I; C3I FFRDC
6	Center for Advanced Aviation System Development		CAASD
7	Center for Enterprise Modernization	IRS Federally Funded Research and Development Center	IRS FFRDC; Internal Revenue Service
8	Center for Naval Analyses		
9	Center for Nuclear Waste Regulatory Analyses		
10	Center for Communications and Computing	Institute for Defense Analyses Communications and Computing	IDA Communications and Computing
11	CMS Alliance to Modernize Healthcare	Centers for Medicare and Medicaid Services Federally Funded Research and Development Center	CMS FFRDC
12	Fermi National Accelerator Laboratory		Fermilab
13	Homeland Security Operational Analysis Center		HSOAC
14	Homeland Security Systems Engineering and Development Institute	Homeland Security Studies and Analysis Institute; Homeland Security Institute	HSSEDI
15	Idaho National Laboratory	Idaho National Engineering and Environmental Laboratory; Idaho National Engineering Laboratory;	

		National Reactor Testing Station	
16	Jet Propulsion Laboratory		JPL
17	Lawrence Berkeley National Laboratory		Lawrence Berkeley Laboratory
18	Lawrence Livermore National Laboratory		Lawrence Livermore Laboratory
19	Lincoln Laboratory		
20	Los Alamos National Laboratory		
21	National Biodefense Analysis and Countermeasures Center		
22	Frederick National Laboratory for Cancer Research	National Cancer Institute at Frederick; Frederick Cancer Research and Development Center; NCI Frederick Cancer Research and Development Center	NCI-Frederick; National Cancer Institute-Frederick; NCI Frederick; National Cancer Institute Frederick
23	National Cybersecurity Center of Excellence		NCCoE
24	National Center for Atmospheric Research		
25	National Defense Research Institute		
26	NSF's National Optical-Infrared Astronomy Research Laboratory	National Optical Astronomy Observatory; Cerro Tololo Inter-American Observatory; Kitt Peak National Observatory; Sacramento	National Optical-Infrared Astronomy Research Laboratory; National Optical

		Peak Observatory	Infrared Astronomy Research Laboratory; NOIRLab
27	National Radio Astronomy Observatory		
28	National Renewable Energy Laboratory	Solar Energy Research Institute	
29	National Solar Observatory		
30	Oak Ridge National Laboratory	Holifield National Laboratory	
31	Pacific Northwest National Laboratory		
32	Princeton Plasma Physics Laboratory		
33	Project Air Force		
34	Sandia National Laboratories		Sandia
35	Savannah River National Laboratory	Savannah River Technology Center	
36	Science and Technology Policy Institute	Critical Technologies Institute	IDA Science and Technology Policy Institute; IDA STPI
37	SLAC National Accelerator Laboratory	Stanford Linear Accelerator Center	
38	Software Engineering Institute		
39	Systems and Analyses Center		IDA SAC
40	Thomas Jefferson National Accelerator Facility	Continuous Electron Beam Accelerator Facility	Jefferson Lab

2) Identifying scientific papers of government laboratories in general

We start by matching based on keywords that often appear in the names of government laboratories. To avoid confounding or matching errors, we first used keywords consisting of two or more consecutive words. Then, for several prolific laboratories with an abbreviated name and one-word keywords representing a government organization, we match the keywords with a strict restriction applied for the string before and after the word. Finally, we used a list of government laboratory names that were collected through a manual search.

Table A1-2. Examples of words included for each of the steps to identify government laboratories in general

Two-word Keywords	'AMERICAN OBSERVATORY', 'METROPOLITAN INSTITUTION', 'METROPOLITAN INST.', 'NATIONAL ACCELERATOR', 'NATIONAL CENTER', 'NATIONAL CYBERSECURITY', 'NATIONAL FACILITY', 'NATIONAL INST.', 'NATIONAL LABORATORIES', 'NATIONAL LABORATORY', 'NATIONAL OBSERVATORIES', 'FEDERALLY FUNDED', 'NATIONAL OBSERVATORY', 'NATIONAL PROGRAM', 'NATIONAL RESEARCH COUNCIL', 'NATIONAL SECURITY', 'NATIONALLY FUNDED', 'POLICY INSTITUTE', 'POLICY RESEARCH'
One-word keywords (inc. abbreviations)	'AGENCY', 'AIR FORCE', 'AMES', 'ARGONNE', 'ARROYO', 'BROOKHAVEN', 'C3I', 'CMS', 'CNRS', 'CSIR', 'INRIA', 'INSERM', 'LAWRENCE', 'LIVERMORE', 'MINISTRY', 'NCATS', 'NCCIH', 'NCMRR', 'NCRR', 'NHGRI', 'NHLBI', 'NIAAA', 'NIAID', 'NIAMS', 'NIBIB', 'NICHD', 'NIDA', 'NIDCD', 'NIDCR', 'NIDDK', 'NIEHS', 'NIGMS', 'NIMH', 'NIMHD', 'NINDS', 'NINR', 'RIKEN', 'SANDIA', 'SAVANNAH', 'TRIUMF'
Manual collection	'ACADEMIA SINICA', 'AGENCY FOR DEFENSE DEVELOPMENT', 'AKADEMIE DER WISSENSCHAFTEN DER DDR', 'AMERICAN HEALTH FOUNDATION', 'AMERICAN RED CROSS', 'CARNEGIE INSTITUTION OF WASHINGTON', 'CENTER FOR ADVANCED AVIATION SYSTEM DEVELOPMENT', 'CENTER FOR COMMUNICATIONS AND COMPUTING', 'CENTER FOR ENTERPRISE MODERNIZATION', 'CENTER FOR NAVAL ANALYSES', 'CENTER FOR NUCLEAR WASTE REGULATORY ANALYSES', 'CENTRE NATIONAL DE LA RECHERCHE

	SCIENTIFIQUE', 'CHINESE ACADEMY OF MEDICAL SCIENCES', 'CMS ALLIANCE TO MODERNIZE HEALTHCARE', 'COMMONWEALTH OF AUSTRALIA', 'COMMONWEALTH SCIENTIFIC & INDUSTRIAL RESEARCH ORGANISATION', ''CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS', 'CONSIGLIO NAZIONALE DELLE RICERCHE', 'COUNCIL OF SCIENTIFIC & INDUSTRIAL RESEARCH', 'DEUTSCHES KREBSFORSCHUNGSZENTRUM', 'HEALTH PROTECTION AGENCY', 'HOMELAND SECURITY OPERATIONAL ANALYSIS CENTER', 'HOMELAND SECURITY SYSTEMS ENGINEERING AND DEVELOPMENT INSTITUTE', 'INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE', 'JAPAN AEROSPACE EXPLORATION AGENCY', 'JAPAN SCIENCE AND TECHNOLOGY AGENCY', 'JET PROPULSION LABORATORY', 'JUDICIARY ENGINEERING AND MODERNIZATION CENTER', 'LINCOLN LABORATORY', 'MAX PLANCK', 'NATIONAL BIODEFENSE ANALYSIS AND COUNTERMEASURES CENTER', 'NATIONAL CANCER INSTITUTE', 'NATIONAL EYE INSTITUTE', 'RESEARCH CENTER BORSTEL', 'SOFTWARE ENGINEERING INSTITUTE'

Note. The complete list is available upon request

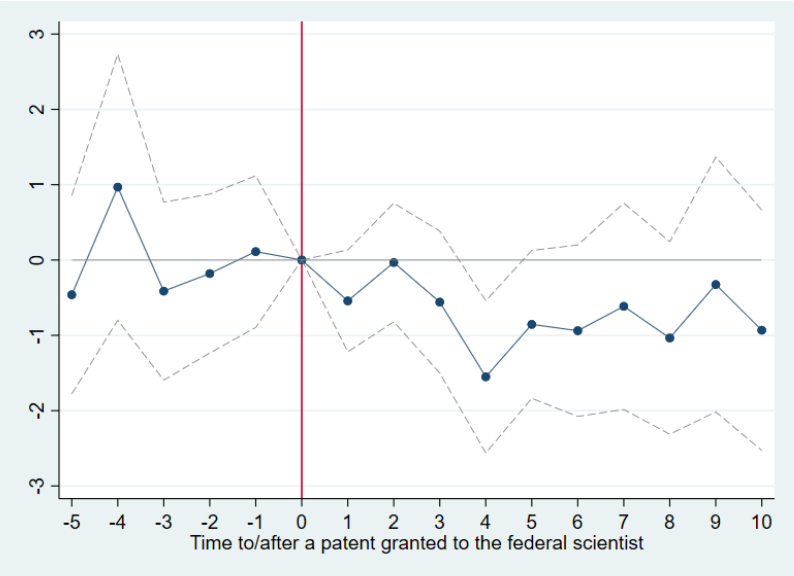
3) Identifying scientific papers of industrial laboratories

We resort to corporate endings of different forms of businesses or their abbreviations. We use the followings: 'CORPORATION', 'CORP.', 'CORP', 'COMPANY', 'INCORPORATED', 'INC.', 'INC', 'LIMITED', 'LTD.', 'LTD', 'GMBH', 'S.P.A.', 'PLC.', 'PLC', 'CO.', 'LLC'. We match these endings on the entire string with a strict restriction applied for the string before and after the word.

Dynamics of the effects of patenting by federal scientist on follow-on inventions

Figure A1 shows the dynamics of the effects of patenting by federal scientists on the rate of follow-on inventions with overlapping subgroups (Panel A) and with no overlapping subgroups (Panel B). We used a ten-year window for the post-period sample to show how long the patent effect lasts for follow-on inventions.

Panel A. Follow-on patents with overlapping subgroups



Panel B. Follow-on patents with no overlapping subgroups

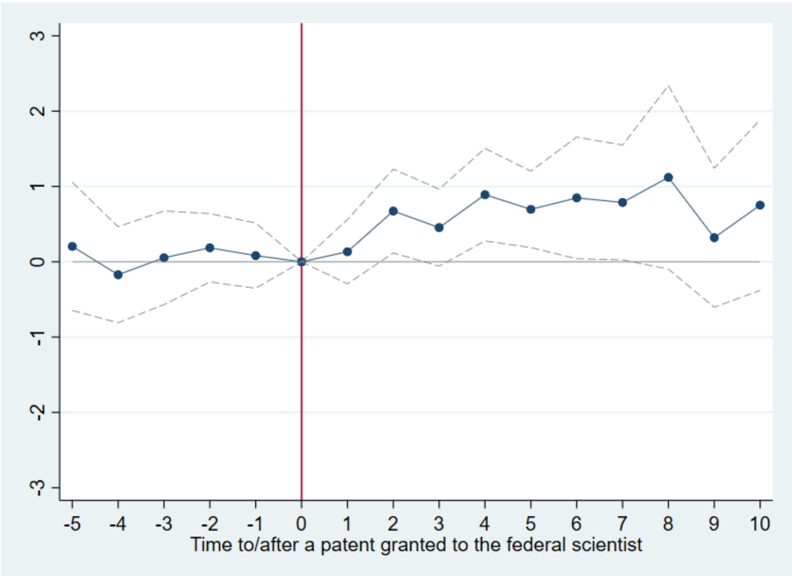


Figure A1-1. Effect of patenting by federal scientist on follow-on inventions (up to 10 years after the focal patent granted)

Appendix 2: Appendix for Chapter four

Matching between surnames in patent and Census data

Matching surnames between Census and patent data requires cleaning of the surname raw strings. We convert all surnames to lower cases and delete unnecessary punctuations and other noise in the surnames (e.g., ’ ”_ / & ; ( ) - =). We also remove suffixes and other extra words after commas (e.g., ‘Foster’, ‘Sr.’, ‘deceased’). This process reduces unique surname strings down to 3,313,643 unique surnames in the Census data and 330,098 unique surnames in the patent data. Out of 374,988 inventor surname raw strings, a total of 275,849 (73.6%) find a match in the census surname. Compared to the matching without these cleaning processes, which finds 230,421 census surname matches out of 374,988 inventor surname raw strings (61.4%), our name cleaning process adds 12.2% of matches. In our data sample specifically, out of 3,165,207 unique inventors that applied for at least one patent in US, 2,894,917 inventors (91.5%) match their surname to the Census data.

Disambiguating geographic location and matching to a county

Although most U.S. patent front pages provide strings for the hometown and state of each inventor, much work must be done to accurately map those strings to counties. Figure A1 illustrates the geographic disambiguation process. We begin with updated data processed via Balsmeier et al. (2018) methods, from 1976 to 2018, which includes 16,215,831 “patent-inventor pairs” because many inventors have multiple patents. Exclusion of non-U.S. and entirely missing data fields leaves 8,065,290 U.S. patent-inventor data points. Amongst these there are 72,122 unique city-state pair strings. Note that this number includes misspellings, neighborhoods and unincorporated areas that do not correspond to city and state, and outright errors.

We exactly matched 27,299 city-state data points for 7,718,350 patent-inventors using the SimpleMaps (https://simplemaps.com/) concordance. We took the remaining unique and unmatched locations and ran them through the Google Geocoding API (https://developers.google.com/maps/documentation/geocoding/overview). This left 10,413 unique city-state pairs and 85,046 patent-inventor pairs, which manual inspection revealed to be mainly errors. 7,980,244 patent inventor pairs were ultimately matched to a city and state, for a 98.9% match rate.

Given that our instrumentation and analysis is at the county level, we need to next map city-state locations to counties. This is complicated by the fact that our data span 1940-2018 and that there have been minor changes to this mapping over time. To address this, we begin with U.S. census records of changes from 1970 to present: (https://www.census.gov/programs-surveys/geography/technical-documentation/county-changes.2010.html). Then, we manually search for changes between 1940 and 1969. We incorporate substantial changes to counties such as county consolidation, part annexation, and FIPS code changes. We build a transitive association file which tracks the changes and anchors all historic changes to the 2020 SimpleMaps concordance (file will be posted upon publication). The 1940 Census doesn't cover VI (Virgin Islands), PR (Puerto Rico), AK (Alaska), and HI (Hawaii), hence, these locations are not considered in the analyses.

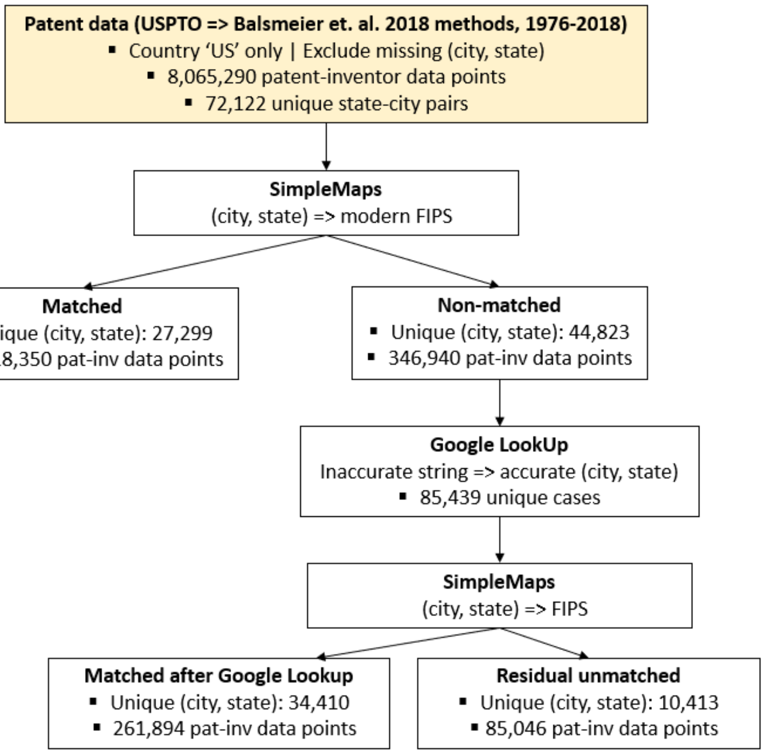


Figure A2-1. Geographic disambiguation process for U.S. inventor city and state

Mapping patent classes and inventors to VentureXpert categories

To estimate the impact of the influx of technology specific inventors on the startup activities of their corresponding industry, we matched NBER technological categories provided by Hall et al. (2001) with VentureXpert industry categories. Table A1 details the manual mapping of NBER technological categories to VentureXpert's major industry groups, i.e., Biotechnology, Communications and Media, Computer Related, Medical/Health/Life Science, Semiconductors/Other Elect, and Non-High-Technology. As underlying technologies overlap between the Biotechnology and Medical/Health/Life Science industry groups, we merged the two industry groups. As VentureXpert does not have corresponding industry groups for mechanical and chemical NBER technological categories, we excluded patent classes corresponding to these technological categories.

Using the concordance between VentureXpert industry groups and NBER patent classification, we classified inventors into each of the five industry groups based on the most frequent industry group that each inventor had patented in. In case of a tie, we took the earliest industry group. We excluded inventors who patented only in patent classes without a corresponding VentureXpert industry group. As a result, out of 763,715 U.S. inventors who had more than two granted patents (whose mobility could be tracked), we were able to assign 602,971 inventors to each of the five VentureXpert industry groups.

Table A2-1. Concordance between VentureXpert industry groups and NBER patent classification

Industry (VentureXpert)	Sub-Category Code	Sub-Category Name	Patent Classes
Biotechnology + Medical/Health/Life Science	31	Drugs	424, 514
	32	Surgery & Medical Instruments	128, 600, 601, 602, 604, 606, 607
	33	Biotechnology	435, 800
	39	Miscellaneous-Drug & Med	351, 433, 623
Communications and Media	21	Communications	178, 333, 340, 342, 343, 358, 367, 370, 375, 379, 385, 455
Computer Related	22	Computer Hardware & Software	341, 380, 382, 395, 700, 701, 702, 704, 705, 706, 707, 708, 709, 710, 712, 713, 714
	23	Computer Peripherals	345, 347
	24	Information Storage	360, 365, 369, 711
Semiconductors/Other Elect	41	Electrical Devices	174, 200, 327, 329, 330, 331, 332, 334, 335, 336, 337, 338, 392, 439
	42	Electrical Lighting	313, 314, 315, 362, 372, 445
	43	Measuring & Testing	73, 324, 356, 374
	44	Nuclear & X-Rays	250, 376, 378
	45	Power Systems	60, 136, 290, 310, 318, 320, 322, 323, 361, 363, 388, 429
	46	Semiconductor Devices	257, 326, 438, 505
	49	Miscellaneous-Elec	191, 218, 219, 307, 346, 348, 377, 381, 386
Non-High-Technology	61	Agriculture, Husbandary, Food	43, 47, 56, 99, 111, 119, 131, 426, 449, 452, 460
	62	Amusement Devices	273, 446, 463, 472, 473

	63	Apparel & Textile	2, 12, 24, 26, 28, 36, 38, 57, 66, 68, 69, 79, 87, 112, 139, 223, 450
	64	Earth Working & Wells	37, 166, 171, 172, 175, 299, 405, 507
	65	Furniture, House Fixtures	4, 5, 30, 70, 132, 182, 211, 256, 297, 312
	66	Heating	110, 122, 126, 165, 237, 373, 431, 432
	67	Pipes & Joints	138, 277, 285, 403
	68	Receptacles	53, 206, 215, 217, 220, 224, 229, 232, 383
	69	Miscellaneous Others	1, 14, 15, 27, 33, 40, 52, 54, 59, 62, 63, 84, 101, 108, 109, 116, 134, 135, 137, 150, 160, 168, 169, 177, 181, 186, 190, 199, 231, 236, 245, 248, 249, 269, 276, 278, 279, 281, 283, 289, 292, 300, 368, 404, 412, 428, 434, 441, 462, 503

Abstract (Korean)

본 학위논문은 과학지식의 확산과 발명가 이동의 관점에서 기술혁신과 지역 경제 성장을 연구한다. 세부적으로, 정부출연연구소의 발명에 대한 특허화가 과학 지식의 확산에 미치는 영향, 발명가의 지리적 이동을 결정짓는 요인, 발명가의 유입이 지역 혁신생태계와 창업 활동에 미치는 영향에 대해 탐구하는 세 가지 소 연구로 구성된다. 각 소 연구의 분석을 위해 대량의 데이터를 수집 및 분석하는 빅데이터 기법과 관찰 데이터에 기반하여 주요 변인 간의 인과 관계를 추론하는 인과추론 기법을 개발 및 적용한다. 각 연구 질문에 대한 이론적 논의와 실증 분석 결과에 기반하여, 기술혁신에 대한 학술적 논의를 확장하고 경영적·정책적 측면의 함의를 제시한다.

첫 번째 소 연구에서는 정부출연연구소의 과학자들의 발명을 특허화 하는 것이 정부출연연구소의 과학지식 확산에 미치는 영향을 규명한다. 기존의 정부출연연구소, 특허보호, 그리고 과학기술 간의 관계에 대한 연구문헌에 기반하여 정부출연연구소 내 과학발명에 대한 특허화가 과학지식 확산에 미칠 수 있는 양면적 영향에 대해 논증한다. 실증분석을 위해 1986년부터 2013년까지 과학 저널에 게재된 전 미국 연방 정부출연연구소의 과학 발명 데이터를 수집한다. Coarsened Exact Matching 방식을 활용하여 관측가능한 특성이 유사한 과학 발명을 추리고 이중차분법(Difference-in-Differences Method)을 통해 발명 특허화의 영향을 비교분석 한다. 이를 통해, 정부출연연구소 과학자가 과학발명에 대해 특허화 하는 것이 같은 기술

분야에서의 혁신은 줄이는 반면 타 기술 분야에서의 혁신은 촉진하는 것을 보인다. 또, 이로 인해 증가하는 기술 혁신은 점진적 발명보다는 위험감수형 발명의 성격을 보이며, 높은 수준의 참신성을 가짐을 확인한다. 특히, 정부출연연구소 연구자들의 특허 등록은 지리적, 기술적으로 거리가 먼 발명가들의 발명활동에 유의한 영향을 미치는 것을 확인한다. 나아가, 특허화의 영향은 해당 연구에 참여한 과학자가 더 적은 사회적 연결성을 가질수록, 해당 과학 지식 분야가 산업에서 덜 익숙한 분야 일수록 더욱 강화되는 것을 확인한다.

첫 번째 소 연구을 통해 정부출연연구소의 과학자들이 직접 참여한 연구에 대해 특허를 출원 및 등록하도록 장려하는 것이 정부출연연구소 지식의 확산과 활용에 긍정적으로 작용한다는 이론적·실증적 근거를 제시한다. 또, 다양한 세부 특성 및 이질성에 대한 검증 결과를 제시함으로써, 정책입안가들이 정부출연연구소의 특허 정책을 수립하는데 활용할 수 있는 실질적 근거를 제시한다. 해당 소 연구의 결과는 정부출연연구소의 지식 확산을 위해 정부출연연구소 과학자들이 외부 기업과 보다 활발히 교류할 수 있는 기회를 만들어야 함을 제시한다. 나아가, 정부출연연구소 내부 과학지식의 확산을 촉진하는 유의한 제도적 방안을 제안함으로써, 정부출연연구소와 산업 내 기업 연구소가 함께 공생하고 상호 보완하는 구조를 만드는데 기여한다.

두 번째 소 연구에서는 발명가의 지역적 이동을 결정하는 인구통계학적 요인에 대해 연구한다. 실증 분석을 위해 1990년과 2010년 사이 미국

내에서 미국 특허를 출원한 발명가 55만여 명의 지리적 이동 내역을 추적하고, 미국 내 3천여개의 자치주(County)별 과거 인구 분포 정보를 수집한다. 변인 간 관계의 분석을 위해 고정효과 회귀분석을 활용한다. 먼저 인구 분포의 측면을 고려하여, 발명가가 다른 지역으로 이동하는 것은 이전 지역에 과거 같은 성씨가 분포되어 있던 정도와 부정적인 관계를 가짐을 확인한다. 이와 같은 과거 성씨 분포의 영향은 발명 경험, 생산성, 질, 이전 이동 패턴, 주 활동 분야 등 발명가 개인 수준의 특성과 현재 위치한 지역의 지역적 특성을 통제한 후에도 여전히 유의한 것을 확인한다. 추가적으로, 성씨의 영향을 조절하거나 영향이 유의하지 않도록 하는 개인적, 가족적, 제도적 요인을 검증한다. 또, 발명가의 성별에 따른 이질성 검증을 통해 성씨 분포의 영향이 성별에 따라 다르게 나타나는 점을 확인한다.

두 번째 소 연구를 통해 과거 성씨의 분포가 발명가들의 지리적 이동에 유의한 영향을 미치는 것을 확인하고 여러가지 다른 요인과의 상호 작용을 살펴본다. 이로써, 발명가의 유입 및 보존을 돕는 요인에 대한 실증적 근거와 함의를 제공한다. 특히, 개인 수준의 발명 관련 요인에 대해 과거 성씨 분포의 영향이 유의하게 변하지 않음을 보여 과거 성씨 분포를 활용한 발명가의 지리적 이동 추적의 범용성을 높인다. 해당 소 연구의 분석 결과는 과거 성씨 분포를 활용하여 Shift-share 도구 변수를 만드는데 대한 타당성을 높인다.

세 번째 소 연구에서는 발명가들의 유입이 해당 지역 내 벤처케피탈의 투자를 받는 신생기업의 창업, 성장 패턴, 그리고 투자 자본 이동에 미치는

영향을 규명한다. 실증 분석을 위해 1987년부터 2007년 사이 벤처케피탈 투자를 받은 미국 전역의 창업기업에 대한 정보를 수집하고 인수합병, 기업공개, 투자유치 등 각 기업별 성장패턴에 대한 정보를 확보한다. 발명가 유입이 지역 창업생태계에 미치는 인과적 영향을 추정하기 위해 미국 내 성씨의 과거 지역분포에 기반한 Shift-share 도구변수를 개발하고 활용한다. 이를 통해, 발명가들의 유입은 해당 지역 내 벤처케피탈의 투자를 받는 신생기업의 창업 수를 늘려주는 것을 확인한다. 특히, 유입되는 발명가들의 기술 분야와 같은 분야에서 신생기업의 창업이 늘어나며 다른 분야에서는 줄어드는 것을 확인하여, 발명가들의 유입에 따라 분야 간의 투자 이동이 일어나는 것을 보인다. 발명가의 유입은 창업의 수뿐만 아니라 인수합병이나 기업공개 등을 통해 성공적으로 투자를 회수하는 창업 기업의 수를 늘려주며, 실패로 이어지는 창업 기업의 수는 줄여주어 창업 기업의 질적 향상에 기여함을 보인다. 나아가, 발명가의 이동은 비첨단 과학 산업의 실패 기업을 줄이고 첨단 과학 산업의 성공 기업을 늘려 벤처케피탈의 자본 투자의 효율을 높임을 보인다.

세 번째 소 연구를 통해 발명가 인적 자본이 신생기업 창업 활동과 투자 자본 이동에 미치는 영향에 대한 실증적 근거를 제시한다. 이는 지역 내 신생기업 창업 활동을 장려하고 자본 투자를 촉진하고자 하는 정책입안가들에게 발명가를 유입하여 창업의 양뿐만 아니라 질, 투자의 효율을 높일 수 있다는 정책적 함의를 제시한다. 추가적으로 해당 소 연구에서 제안한 도구변수는 향후 다양한 인적자본 이동이 지역 단위에

미치는 영향을 규명하는데 유용하게 활용될 것으로 기대한다.

결론적으로 본 학위논문은 과학 지식과 발명가의 이동이 혁신과 경제성장에 미치는 영향에 대한 이해를 높인다. 또, 혁신에 대한 이론적 고찰과 실증적 분석을 통해 실무적으로 적용 가능한 경영적·정책적 함의를 제시한다. 학문적으로도 각 소 연구 별로 관련된 선행연구 흐름을 확장하는 주요한 학문적 기여점을 갖는다.

주요어 : 과학지식, 정부출연연구소, 지식확산, 발명가 이동, 혁신, 기업가정신
학　번 : 2017-31119