Patents, Spillovers, and the Allocation of Talent

Microeconometric Perspectives on the Knowledge Economy

Inaugural-Dissertation

zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2017

vorgelegt von

Markus Nagler

Referentin:	Prof. Dr. Monika Schnitzer
Koreferent:	Prof. Dr. Ludger Wößmann
Promotionsabschlussberatung:	31. Januar 2018

Für meine Familie.

À minha familia.

Acknowledgements

First and foremost, I would like to thank my supervisor Monika Schnitzer for her continuous and outstanding support over the past years. Since the very beginning of my undergraduate studies, she has given me the opportunity to gain insights not only within economics, but also on a professional and on a personal level. I am indebted to her for much of what I take away from these years of study. Ludger Wößmann also played a very important role. He sparked my interest in empirical economics, was pivotal in my research education, and always wholeheartedly supported me. I am very grateful to both for guiding and supporting me throughout the past years. I would also like to thank Oliver Falck for completing my dissertation committee. Since my days as an undergraduate student, he has always been very helpful and supportive.

I am very grateful to Martin Watzinger, without whom this dissertation would never have been possible. His guidance, his patience with my ideas, and his example inspired me to push harder, learn more, and think deeper about economics. I also thank Anna Gumpert, whose thoughtful advice and sincere support were instrumental in many moments during my studies.

My colleagues at the Munich Graduate School of Economics, at the Evidence-Based Economics Program, and especially at the Seminar for Comparative Economics were always helpful and made the years of graduate school much nicer. Chen Li and Henrike Steimer were great office mates. Thank you! Natalie Obergruber, Matthias Wilhelm, and Cathrin Mohr deserve special mention. They helped to make some difficult times less painful and the good times even better. I would also like to thank Karin Fritsch, Julia Zimmermann, Dagmar Ehrhardt, and the student assistants at the Seminar for Comparative Economics for their great help.

My co-authors (in chronological order) Ludger Wößmann, Sascha Becker, Marc Piopiunik, Martin West, Martin Watzinger, Thomas Fackler, Monika Schnitzer, Lukas Treber, Sarah Weise, and Jeffrey Furman deserve special credit for this dissertation. I learned a lot from working with them and would have not been able to realize any of the projects without the insights I gained from our joint work and the discussions.

While writing this dissertation I was lucky to spend extended periods abroad. I would like to thank Martin West for making possible my visit to the Program on Education Policy and Governance at Harvard University, which opened many doors. And I am very grateful to David Autor and the Department of Economics at MIT for their hospitality during the past academic year. The friendly and vibrant atmosphere, the open arms with which I was welcomed, and the rigor of intellectual thought at the department made the year both very enjoyable and exciting. I learned something new every day. My special gratitude goes to my friends and colleagues who I met during this year.

The DAAD allowed me to not only spend a year at MIT, but also to study at UCL before my doctoral studies. I am very grateful for these opportunities. I would also like to thank the German Science Foundation and the EliteNetwork of Bavaria for their financial

support. This dissertation also benefited from numerous comments and discussions at conferences and seminars.

Without my good friends, I would not have been able to achieve this. And more importantly, life would have been much less enjoyable. I am very grateful for having you.

Most of all, I thank my family. Without their endless love and support, none of this would have been possible.

<div align="right">
Markus Nagler
September 2017
</div>

Contents

CONTENTS

List of Tables

Contents

List of Figures

Contents

Preface

"Traditionally, wealth was defined by land and natural resources.
Today the most important resources is between our ears."

Barack Obama, 2014

The world today is a world of ideas: knowledge, technology, and human capital are drivers of wealth and economic growth (Jones and Romer, 2010). The increased reliance on these factors as sources of economic prosperity has been labeled as a change towards a "knowledge economy" (e.g., The Economist, 2000; Mokyr, 2002; Powell and Snellman, 2004). Using microeconometric methods, this dissertation sheds light on the underlying mechanisms behind two central elements of the knowledge economy. The first part focuses on the role of intellectual property rights for the diffusion of ideas. The second part investigates productivity determinants of "knowledge workers".[1] Both parts consist of two self-contained chapters each.

The first part of the dissertation provides evidence on how patents shape cumulative innovation. They are motivated by the insight that almost all new knowledge builds on prior ideas (Scotchmer, 1991). Most famously, this was described by Isaac Newton as "standing on the shoulders of giants". This feature of the innovation process is central to theories of endogenous growth (e.g., Romer, 1986, 1990, 1994; Jones, 1995; Weitzman, 1998). Therefore, understanding barriers to knowledge diffusion is crucial. One determinant that has received substantial interest in recent years is the patent system. Due to the public goods nature of innovation, governments around the world rely on patents to provide incentives

---

[1]While there is no unique definition of this expression, knowledge workers are thought to be those workers whose jobs mainly consist of using and/or generating knowledge and information (e.g., Drucker, 1999; The Economist, 2005; Wall Street Journal, 2016).

for research and development of new technologies. The classical viewpoint in economics is that patents should ideally solve the trade-off between incentives for innovation, disclosure of ideas, and dead-weight loss due to monopoly power in the period during which the patent is valid. However, there is evidence that patents impose additional costs on innovation: in particular, the debate about the *dynamic* impacts of intellectual property rights has questioned the benefits of patents (Hall and Harhoff, 2012). An emerging literature has shown that intellectual property rights may dampen cumulative innovation (Williams, 2013; Galasso and Schankerman, 2015, forthcoming; Sampat and Williams, 2015). This is a potentially important drawback of patents.

The first chapter, which is based on joint work with Martin Watzinger, Thomas Fackler, and Monika Schnitzer (Watzinger et al., 2017a), builds on this literature. It investigates whether patents held by a dominant firm are harmful for follow-on innovation, and if so, whether antitrust enforcement in the form of compulsory licensing of patents provides an effective remedy. We advance on these questions by analyzing the effects of one of the most important antitrust rulings in U.S. history: the 1956 consent decree against the Bell System. This decree settled a seven-year antitrust lawsuit that sought to break up the Bell System, which was charged with having foreclosed competitors from the market for telecommunications equipment. The decree forced Bell to license all its 7'820 patents royalty-free and the company was barred from being active in any industry other than telecommunications. At that time, it was one of the most innovative companies in the world, with their research subsidiary Bell Labs producing path-breaking innovations in a broad range of technologies.

Our analysis shows that compulsory licensing increased innovation that built on Bell patents. We find that in the first five years after the decree, follow-on innovation increased by 17% or a total of around 1'000 citations. Additionally, the number of patents increased in fields with compulsorily licensed patents compared to similar fields without. These effects are mainly driven by young and small companies. The positive effects of compulsory licensing were however restricted to industries other than the telecommunications equipment industry. Thus, compulsory licensing without structural remedies appears to be an ineffective remedy for market foreclosure.

We contribute to the literature on intellectual property by providing robust causal evidence for the negative effects of patents on follow-on innovation, especially of small and young companies. We examine both short-run direct effects as well as long-run effects in a unique set-up that enables us to address key challenges for the impact evaluation of compulsory licensing. Our finding of company entries as the main mechanism driving its positive innovation effects is consistent with concerns that patents may prevent new and innovative firms from entering markets.

We are also the first to empirically investigate the effect of antitrust enforcement on innovation. Overall, our results suggest that market foreclosure slows down technological progress and that antitrust enforcement can have a lasting positive impact on the long-run rate of technological change if market entry is not hindered by exclusionary practices. Yet, compulsory licensing without structural remedies is not sufficient to overcome foreclosure. This chapter finally contributes to our understanding of innovation and growth in the United States in the twentieth century. By providing free state-of-the-art technology to all U.S. companies, compulsory licensing increased U.S. innovation because it opened up new markets for a large number of entrants.

The way in which patents affect cumulative innovation is more nuanced, however. To balance monopoly rights, patents require the disclosure of underlying technical information. This is deemed one of the main advantages of the patent system as inventors can build on previous ideas (e.g., Machlup, 1958; Hall and Harhoff, 2012). This feature is therefore crucial when analyzing dynamic effects of patents on cumulative innovation. Yet, the question whether this aspect is indeed important is particularly challenging to analyze empirically (Graham and Hegde, 2015; Hegde and Luo, 2017; Williams, 2017). Since the patent system makes monopoly rights dependent on the disclosure of technical information, there is little variation allowing identification of the "enablement effect" of disclosure separately from the effects of exclusion. Therefore, it is hard to judge whether the "grand bargain" in the patent system is a profitable one for society.

The second chapter, which builds on joint work with Jeffrey Furman and Martin Watzinger (Furman et al., 2017), takes advantage of the expansion of the U.S. Patent and Trademark Depository Library Program from 1977 to 1997 to investigate the effects of disclosure of

patent information on innovation. This patent library system was created in the late 1800s to provide patents and innovation-related resources for independent inventors, entrepreneurs, and incumbent firms. While the exclusion rights associated with patents are national in scope, the opening of these patent libraries in a period before the Internet yielded regional variation in the costs to access the technical information disclosed in patent documents.

By comparing the number of patents around patent libraries to their number around control libraries before and after the opening, we can thus identify the impact of disclosure on follow-on innovation. As control group, we use regions around Federal Depository Libraries (FDLs). As the missions of patent libraries and FDLs are similar, namely to provide the public with official documents, almost all patent libraries are also Federal Depository Libraries. Therefore, these libraries are a natural comparison group to patent libraries.

We find that after a patent library opens the number of patents around the library increases by around two patents per year on average or 18% relative to the pre-opening mean. In line with increased access to patents driving this effect, we find that young and small companies increase patenting more. These inventors plausibly face larger barriers to access than large companies. In additional analyses, we also find that the structure of patents changes after a patent library opens: the distance to cited patents increases for new applications by inventors close to a patent library. Inventors therefore start to work on problems that are less local, rendering the geography of innovation more dispersed.

This study is the first to show that access to technical information disclosed in patents can increase innovation. Disclosure is often taken as one justification for the monopoly rights attached to patents. Yet, critics often question its usefulness (e.g., Roin, 2005; Gambardella et al., 2011). Our study adds to the literature in showing that access to technical information provided by patent libraries increases patenting for small and young companies. More generally, our study contributes to the literature on research enhancing institutions (e.g., Furman and Stern, 2011). These institutions lower the costs of access to useful knowledge and thus help to foster geographical and intertemporal spillovers, ultimately fueling economic growth (Mokyr, 2002). We contribute by showing that patent

libraries increased innovation across U.S. states by improving access to technical information.

The second part of the dissertation turns to the productivity of knowledge workers. These professions, which perform non-routine, non-manual tasks and require high cognitive skills, became much more important in developed economies in the past decades (e.g., Neef, 1998; Acemoglu and Autor, 2011; Autor, 2014).[2] For example, since the 1960s, the number of workers in science and engineering has grown substantially faster than the total workforce in the United States (National Science Board, 2016). In the business literature, raising knowledge workers' productivity has even been called the "most important contribution of management in the 21st century" (Drucker, 1999, p. 79). For economists, interesting aspects about knowledge workers are the abstract nature and unspecific goal of their tasks, the difficulty of monitoring effort, and the importance of *other* knowledge workers for their productivity (e.g., Manso, 2011; Catalini, forthcoming). Also, knowledge workers may be different from the general population in terms of job preferences. For jobs in teaching or science, intrinsic motivation and occupational selection may be more important for productivity than for blue-collar workers (e.g., Stern, 2004; Besley and Ghatak, 2005). Therefore, empirically studying determinants of knowledge workers' productivity is an important endeavor of economic research. To study these determinants, this dissertation turns to occupations which are fundamental for knowledge production: scientists and teachers. The two chapters emphasize the importance of the geographic and the sectoral allocation of talent for knowledge workers' productivity.

The third chapter asks whether labor mobility is an important determinant of inventors' productivity. One crucial input in the ideas production function are the ideas of *other* scientists (e.g., Weitzman, 1998). Local access to ideas may therefore be an important driver of individual productivity. After all, as Glaeser et al. (1992) write, "intellectual breakthroughs must cross hallways and streets more easily than oceans and continents" (p.1127). This is increasingly important for scientists, given the relevance of recombination for truly novel academic ideas (e.g., Uzzi et al., 2013) and the growing importance of teams

---

[2]Note that while the demand for cognitive skills has apparently slowed during the 2000s, the reasons for this are still unclear (Autor, 2015; Beaudry et al., 2016; Deming, forthcoming).

in the production of new knowledge (e.g., Wuchty et al., 2007; Jones, 2009; Stephan, 2012). In the innovation literature, geography is therefore regarded as a "key factor in explaining the determinants of innovation and technological change" (cf. Audretsch and Feldman, 2004, p.2715).

We analyze the impact of scientists' mobility on their academic output. Mobility is often mentioned as a way how to improve access to localized knowledge spillovers. Yet, little is known about what its effects really are. The econometric challenge when estimating impacts of mobility is that scientists self-select into moving. In particular, scientists likely move to places where they are more productive. Any observed correlation between labor mobility and scientific productivity might therefore be due to other factors influencing both.

To circumvent this identification problem, we exploit an institutional feature of the German university system: hiring committees are required to create a short list of suitable candidates for each appointment.[3] We have access to these lists for the years 1950-2005 from one large university in Germany that offers a wide range of fields. In this chapter, we use non-moving candidates on these ranked lists as counterfactuals for the moving scientist. This setup provides two main advantages: on the one hand, it circumvents the problem of selection into moving as all scientists on the list showed interest in moving to the destination university. On the other hand, candidates on the appointment list are qualitatively comparable for institutional reasons. Therefore, non-moving scientists on the same appointment list provide a credible estimate of what would have happened had the moving scientist not been appointed.

We find that after a move, a scientist's productivity as measured by quality-weighted publications increases by around 13% relative to the control group of non-moving scientists. In contrast, there is no difference in academic output between movers and non-movers before the move or between higher- and lower-ranked non-movers. The results are entirely driven by scientists in the natural sciences and by those from lists with an above-average number of citations to pre-move work. We provide evidence suggesting that an alternative explanation of the effects as returns to lab ownership or increased funding is unlikely.

---

[3]The setup and identification strategy of this paper closely tracks the companion paper, Watzinger et al. (2017b).

Overall, we think the estimates are consistent with the idea that access to local knowledge and possibilities of recombination increase in response to the move.

Economists have long attempted to understand the driving factors behind the productivity of scientists (e.g., Arrow, 1962). This study is one of the first to rigorously estimate the impact of labor mobility for the productivity of (academic) researchers. Little is known about this impact even though most science systems in the world incorporate features which increase labor mobility. The previous literature has relied on matching strategies and debatable instruments. In contrast, our setup of scientists applying for the *same position* at the *same university* provides a credible counterfactual for movers. What is more, this is the first analysis to assess the heterogeneity of impacts of labor mobility on innovation across high- and low-impact scientists and across different fields. Our results suggest that labor mobility may indeed be a fruitful way to increase academic productivity for scientists.

The final chapter, which is based on joint work with Marc Piopiunik and Martin West (Nagler et al., 2015), focuses on the *sectoral* allocation of talent. More specifically, we analyze to which extent the selection of talent into careers is affected by the relative compensation in a profession. Changes in the selection into sectors have long been recognized as important for productivity and growth (e.g., Murphy et al., 1991). In recent years, a vibrant literature has documented impacts of macroeconomic conditions on workers who started their careers during times of crises (Kahn, 2010; Oreopoulos et al., 2012). The counterpart to these "scarring effects" is the improvement in the average ability of individuals entering some lower-paying or higher-risk occupations during recessions (e.g., Oyer, 2008; Boehm and Watzinger, 2015; Shu, 2012). This literature has so far analyzed impacts on small groups in the labor market, such as academic economists or MBA students.

We focus on teachers, who are a prime example of knowledge workers and make up around three percent of all U.S. full-time workers. The importance of teachers for the creation of human capital has been widely recognized in the economics literature (e.g., Hanushek, 2011; Hanushek and Rivkin, 2012; Chetty et al., 2014a,b). Their "output", higher human capital among students, explains large parts of income differences between and within countries (e.g., Mankiw et al., 1992; Hanushek and Woessmann, 2008; Jones,

2016; Hanushek et al., forthcoming). Yet, the average salary of teachers is relatively low compared to their required qualifications in many countries around the world, which has often been cited as a key reason why higher-skilled individuals do not want to become teachers (e.g., Dolton and Marcenaro-Gutierrez, 2011). Existing research investigating this link has focused on regional differences in relative pay or on long-run changes in labor market opportunities (e.g., Bacolod, 2007; Britton and Propper, 2016). However, it suffers from two key limitations. First, relative regional pay may be endogenous to teacher effectiveness. Second, widely used observable measures such as academic credentials are poor predictors of actual classroom impact (cf. Jackson et al., 2014).

We exploit macroeconomic conditions at career start as a source of exogenous variation in the outside labor-market options of potential teachers. The idea is that teaching is a relatively stable occupation over the business cycle. In our Roy-style framework (Roy, 1951), more high-ability individuals choose teaching over other professions during recessions because of lower (expected) earnings in those alternative occupations. To measure teacher quality, we construct estimates of teachers' value-added to student test scores, a widely used measure of teacher effectiveness (e.g., Jackson et al., 2014). These estimates are based on administrative data on around 33'000 teachers and their students in the Florida public school system in the school years 2000-01 through 2008-09.

Our results show that teachers who entered the profession during recessions are roughly 0.10 standard deviations more effective in raising math test scores than teachers who entered the profession during non-recessionary periods. As business cycle conditions at career start are exogenous to teacher quality, we interpret our reduced-form estimates as causal effects. We provide evidence that our results reflect changes in the supply of potential teachers rather than demand changes by school districts. Our results have far-reaching consequences: based on figures from Chetty et al. (2014b), the difference in average math teaching effectiveness between recession and non-recession entrants implies a difference in students' discounted life-time earnings of around $13'000 per classroom taught each year. Through a back-of-the-envelope calculation, we argue that it would be economically beneficial to increase teacher pay in Florida.

Magnitudes aside, our findings suggest that policymakers would be able to attract more

effective individuals into the teaching profession by raising the economic benefits of becoming a teacher. This is not a trivial result. If intrinsic motivation positively affects teachers' effectiveness, then increasing teacher pay may at the margin attract less effective individuals into the teaching profession. Since we find the opposite, intrinsic motivation seems to be of second-order importance relative to the effects of increasing teacher pay on selection when hiring more effective teachers. More generally, recessions may provide a window of opportunity for the public sector to hire more able applicants.

Our study is the first to document a causal effect of outside labor-market options on the effectiveness of entering teachers in raising student test scores. In comparison to the previous literature, we advance by using exogenous changes in the relative compensation of teachers, by using a direct and validated measure of teacher quality, and by isolating the *selection into teaching* as opposed to effects of pay on effort or retention.

In summary, this dissertation offers new insights into driving forces behind the determinants of cumulative innovation and the productivity of knowledge workers. These microeconometric perspectives on the knowledge economy may hopefully contribute to designing economic policies that account for the increasing importance of ideas, technology, and human capital for economic prosperity.

Chapter 1

Antitrust, Patents, and Cumulative Innovation: Bell Labs and the 1956 Consent Decree

1.1 Introduction

Innovation is a key driver of economic growth. One of the main instruments governments use to foster innovation is the patent system. A patent gives the right to exclude others from using the patented inventions in order to stimulate innovation. However, there is a growing concern that dominant companies might use patents strategically to deny potential entrants, often small technology-oriented start-ups, access to key technologies in an attempt to foreclose the market.[1] As start-ups are thought to generate more radical innovations than incumbents, market foreclosure may harm technological progress and economic growth (Baker, 2012).[2] To address this problem many critics call for antitrust

---

This chapter is based on joint work with Martin Watzinger, Thomas Fackler, and Monika Schnitzer (Watzinger et al., 2017a).

[1]Derek Thompson, "America's Monopoly Problem", The Atlantic, October 2016; Robert B. Reich, "Big Tech Has Become Way Too Powerful," The New York Times, September 18, 2015, p. SR3; Michael Katz and Carl Shapiro "Breaking up Big Tech Would Harm Consumer," The New York Times, September 28, 2015, p. A24; Thomas Catan "When Patent, Antitrust Worlds Collide," Wall Street Journal, November 14, 2011.

[2]For example, Akcigit and Kerr (forthcoming) show that start-ups do more explorative research and Foster et al. (2006) show that in the retail sector the fast pace of entry and exit is associated with productivity-enhancing creative destruction

policies as a remedy (Wu, 2012; Waller and Sag, 2014). Yet, up to now there are no empirical studies showing that antitrust enforcement can effectively promote innovation.

In this chapter we investigate whether patents held by a dominant firm are harmful for follow-on innovation, and if so, whether antitrust enforcement in the form of compulsory licensing of patents provides an effective remedy. We advance on these questions by analyzing the effects of one of the most important antitrust rulings in U.S. history: The 1956 consent decree against the Bell System. This decree settled a seven-year old antitrust lawsuit that sought to break up the Bell System, the dominant provider of telecommunications services in the U.S., because it allegedly monopolized "the manufacture, distribution, and sale of telephones, telephone apparatus and equipment" (Antitrust Subcommittee, 1958, p.1668). Bell was charged with having foreclosed competitors from the market for telecommunications equipment because its operating companies had exclusive supply contracts with its manufacturing subsidiary Western Electric and because it used exclusionary practices such as the refusal to license its patents.

The consent decree contained two main remedies. The Bell System was obligated to license all its patents royalty-free and it was barred from entering any industry other than telecommunications. As a consequence, 7'820 patents or 1.3% of all unexpired U.S. patents in a wide range of fields became freely available in 1956. Most of these patents covered technologies from the Bell Laboratories (Bell Labs), the research subsidiary of the Bell System, arguably the most innovative industrial laboratory in the world at the time. The Bell Labs produced path-breaking innovations in telecommunications such as the cellular telephone technology or the first transatlantic telephone cable. But more than half of its patents were outside the field of telecommunications because of Bell's part in the war effort in World War II and its commitment to basic science. Researchers at Bell Labs are credited for the invention of the transistor, the solar cell, and the laser, among other things.

The Bell case is uniquely suited to investigate the effects of compulsory licensing as an antitrust measure for two reasons: First, it allows to study the effects of compulsory licensing without any confounding changes in the market structure. In compulsory licensing cases, antitrust authorities usually impose structural remedies such as divestitures, which

makes it difficult to separate the innovation effects from changes in the market structure from the innovation effects from changes in the licensing regime. Yet, in the case of Bell no structural remedies were imposed, despite the original intent of the Department of Justice. This was due to the intense lobbying of the Department of Defense as Bell was considered vital for national defense purposes.

Second, Bell's broad patent portfolio enables us to measure the effect of compulsory licensing on follow-on innovation in different competitive settings. 42% of Bell's patents were related to the telecommunications industry. In this industry, Bell was a vertically integrated monopolist who allegedly foreclosed rivals. The remaining 58% of Bell's patent portfolio had its main application outside of telecommunications. In these industries, Bell was not an active market participant. By looking at the differential effects of compulsory licensing inside and outside of the telecommunications industry we can distinguish the effects of potential foreclosure of patents and of potential bargaining failures that are inherent in the patent system.

Our analysis shows that compulsory licensing increased follow-on innovation that builds on Bell patents. This effect is driven mainly by young and small companies. But the positive effects of compulsory licensing were restricted to industries other than the telecommunications equipment industry. This suggests that Bell continued to foreclose the telecommunications market even after the consent decree took effect. Thus, compulsory licensing without structural remedies appears to be an ineffective remedy for market foreclosure. The increase of follow-on innovation by small and young companies is in line with the hypothesis that patents held by a dominant firm are harmful for innovation because they can act as a barrier to entry for small and young companies who are less able to strike licensing deals than large firms (Lanjouw and Schankerman, 2004; Galasso, 2012). Compulsory licensing removed this barrier in markets outside the telecommunications industry, arguably unintentionally so. This fostered follow-on innovation by young and small companies and contributed to the long run technological progress in the U.S.

Looking at the results in more detail, we first consider the effect of compulsory licensing on innovations that build on Bell patents. We measure follow-on innovation by the number of patent citations Bell Labs patents received from other companies that patent in the

U.S. We find that in the first five years follow-on innovation increased by 17% or a total of around 1'000 citations. Back-of-the-envelope calculations suggest that the additional patents other companies filed as a direct result of the consent decree had a value of up to $5.7 billion in today's dollars. More than two-thirds of the increase is driven by young and small companies and individual inventors unrelated to Bell. Start-ups and individual inventors increase follow-on innovation by 32% while for large and old companies the increase is only around 6%. Robustness checks show that the increase in follow-on innovation is not driven by simultaneous contemporary shocks to technologies in which Bell was active or by citation substitution.

The increase in follow-on innovation by other companies is accompanied by a decrease in follow-on innovation by Bell, but this negative effect is not large enough to dominate the positive effect on patenting by others. The limited negative response by Bell is most likely due to the fact that at the time of the consent decree, Bell was a regulated monopolist subject to rate of return regulation. Yet, the consent decree changed the direction of Bell's research. Bell shifted its research program to focus more on telecommunications research, the only business Bell was allowed to be active in.

In a second step we split the increase in follow-on innovation by industry. We do not find any increase in innovation in the telecommunications industry, the aim of the regulatory intervention. Compulsory licensing fostered innovation only outside of the telecommunications industry. This pattern is consistent with historical records that Bell continued to use exclusionary practices after the consent decree took effect and that these exclusionary practices impeded innovation (Wu, 2012). As no structural remedies were imposed, Bell continued to control not only the production of telephone equipment but was - in the form of the Bell operating companies - also its own customer. This made competing with Bell in the telecommunications equipment market unattractive even after compulsory licensing facilitated access to Bell's technology. For example, the Bell operating companies refused to connect any telephone that was not produced by Western Electric, the manufacturing subsidary of the Bell System (Temin and Galambos, 1987, p.222). In other industries, compulsory licensing was effective to foster innovation by young and small companies since Bell as the supplier of technology did not control the product markets through vertical integration or via exclusive contracts.

Although the 1956 consent decree was not effective in ending market foreclosure, it permanently increased the scale of U.S. innovation. In the first five years alone, the number of patents increased by 25% in fields with compulsorily licensed patents compared to technologically similar fields without; and it continued to increase thereafter. This increase is again driven by small and new companies outside the telecommunications industry. We find only a small increase in patents related to the production of telecommunications equipment. This indicates that market foreclosure may slow down technological progress and suggests that antitrust enforcement can have an impact on the long-run rate of technological change.

We contribute to the literature by being the first to empirically investigate the effect of antitrust enforcement on innovation. Our results suggest that foreclosure impedes innovation and that compulsory licensing without structural remedies is not sufficient to overcome foreclosure. Access to technology through compulsory licensing alone does not stimulate market entry and innovation unless there is sufficient access to the product market as well. These insights are relevant not only for antitrust cases about abuse of a dominant market position, such as the Bell case, but also for merger and acquisition cases where compulsory licensing is often used as a remedy when mergers are approved. Our empirical findings support theoretical arguments in the antitrust literature suggesting that to increase innovation, antitrust measures should focus on exclusionary practices and the protection of start-ups (Segal and Whinston, 2007; Baker, 2012; Wu, 2012).

We also contribute to the literature on intellectual property by providing robust causal evidence for the negative effects of patents on follow-on innovation of small and young companies. Our estimate of an increase in follow-on innovation by 17% is significantly smaller than the increase reported by Galasso and Schankerman (2015). They study the innovation effect of litigated and invalidated patents and find an increase of 50%.[3] While our study looks mainly at patents in the electronics and computer industry, Sampat and Williams (2015) consider gene patents and find no effect on follow-on research. The size of our measured effects is consistent with that reported by other studies such as Murray and Stern (2007) and Moser and Voena (2012). They study various measures of follow-on

---

[3]Litigated patents are selected by importance and by the virtue of having a challenger in court. Thus, the blocking effects of these particular patents might be larger than the average effect for the broad cross-section of patents.

innovation and report an overall impact of a patent removal of about 10-20% in biotech and chemistry. Our finding of entry of companies as the main mechanism driving the positive innovation effects of compulsory licensing is consistent with Galasso and Schankerman (2015). They show that the increase in citations can be attributed to small companies citing invalidated patents of large companies.

Finally, this study contributes to our understanding of innovation and growth in the United States in the twentieth century. By providing free state-of-the-art technology to all U.S. companies, compulsory licensing increased U.S. innovation because it opened up new markets for a large number of entrants. This interpretation is consistent with theoretical concepts and historical accounts. Acemoglu and Akcigit (2012) show theoretically that compulsory licensing can foster innovation because it enables more companies to compete for becoming the leader in an industry.[4] In line with this idea, Gordon Moore, the co-founder of Intel, stated that "One of the most important developments for the commercial semiconductor industry (...) was the antitrust suit filed against [the Bell System] in 1949 (...) which allowed the merchant semiconductor industry "to really get started" in the United States (...) [T]here is a direct connection between the liberal licensing policies of Bell Labs and people such as Gordon Teal leaving Bell Labs to start Texas Instruments and William Shockley doing the same thing to start, with the support of Beckman Instruments, Shockley Semiconductor in Palo Alto. This (...) started the growth of Silicon Valley" (Wessner et al., 2001, p. 86). Similarly, Peter Grindley and David Teece opined that "[AT&T's licensing policy shaped by antitrust policy] remains one of the most unheralded contributions to economic development – possibly far exceeding the Marshall plan in terms of wealth generation it established abroad and in the United States" (Grindley and Teece, 1997).

The remainder of this chapter is organized as follows. Section 1.2 describes the antitrust lawsuit against Bell and the consent decree. In Section 1.3, we describe the data and the empirical strategy. In Section 1.4, we show that compulsory licensing increased follow-on

---

[4]In the model of Acemoglu and Akcigit (2012), compulsory licensing also makes innovation less profitable because leaders are replaced more quickly. In the case of Bell, compulsory licensing was selectively applied to only one company which was not active in the newly created industries. This suggests that there was no disincentive effect and that our empirical set-up cleanly measures the effects of an increase in competition on innovation.

innovation and conduct robustness checks. In Section 1.5, we examine the effectiveness of compulsory licensing as an antitrust measure against foreclosure in the market for telecommunications equipment. In section 1.6, we present the long run effects of the consent decree on U.S. patenting. Section 1.8 concludes.

1.2 The Bell System and the Antitrust Lawsuit

In this section we describe the Bell System and the antitrust lawsuit against Bell. We then discuss the unique features of the case that make it ideally suited for our empirical analysis.

1.2.1 The Bell System was a Vertically Integrated Monopolist

In 1956, American Telephone & Telegraph (AT&T) was the dominant provider of telecommunications services in the U.S. Through its operating companies, it owned or controlled 98% of all the facilities providing long distance telephone services and 85% of all facilities providing short distance telephone services. These operating companies bought all of their equipment from Western Electric, the manufacturing subsidiary of AT&T. As a consequence, Western Electric had a market share in excess of 90% in the production of telecommunications equipment. Western Electric produced telecommunications equipment based on the research done by the Bell Laboratories, the research subsidiary of AT&T and Western Electric. All these companies together were known as the Bell System, stressing its complete vertical integration (Figure 1.1). In terms of assets, AT&T was by far the largest private corporation in the world in 1956, employing 598'000 people with an operating revenue of $ 2.9 billion or 1% of the U.S. GDP at the time (Antitrust Subcommittee, 1959, p.31).

The Bell System held patents on many key technologies in telecommunications, as well as a large number of patents in many other fields. Between 1940 and 1970, Bell filed on average ∼543 patents or 1% of all U.S. patents each year. More than 70% of the patents protected inventions of the Bell Laboratories (Bell Labs), arguably the most innovative industrial laboratories in the world at the time.

Figure 1.1: The Bell System
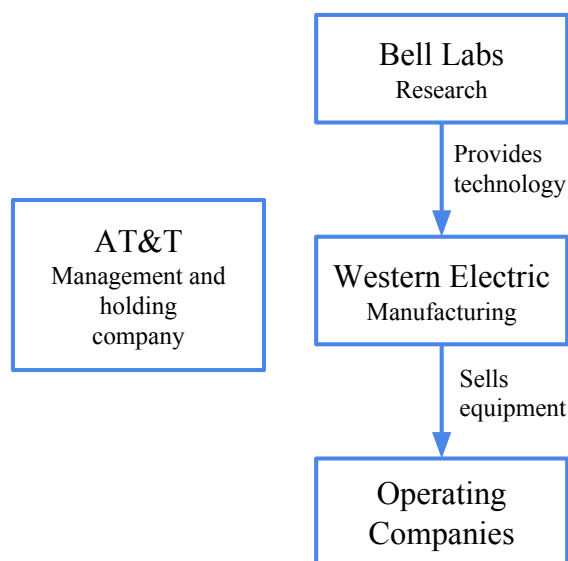
The Bell Labs were unique in their commitment to basic research. When the Bell Labs were founded in 1925, no one knew which part of science might yield insights for the problems of electric communication (Rosenberg, 1990; Nelson, 1962, p.31). As a result, the Bell System decided that - besides supporting the day-to-day need of the System - the Bell Labs would engage in basic science, assuming it would eventually yield products for some part of the large Bell System (Gertner, 2012; Nelson, 1959; Arora et al., forthcoming, p. 31).[5]

The Bell Labs produced path-breaking basic and applied research. Scientists at Bell are credited for the development of radio astronomy (1932), the transistor (1947), cellular telephone technology (1947), information theory (1948), solar cells (1954), the laser (1957), and the Unix operating system (1969). The 1950 staff of Bell Labs alone included four future Nobel Laureates in physics, one Turing Award winner, five future U.S. National Medals of Science recipients and 10 future IEEE Medals of Honor recipients. In 1950, Bell Labs employed 6'000 people, one third of whom were professional scientists and engineers (Nelson, 1962; Temin and Galambos, 1987). This was 1% of the entire science and

---

[5]According to the first head of basic and applied research at Bell Labs, Harold Arnold, his department would include "the field of physical and organical chemistry, of metallurgy, of magnetism, of electrical conduction, of radiation, of electronics, of acoustics, of phonetics, of optics, of mathematics, of mechanics, and even of physiology, of psychology and meteorology". This broad focus led to major advances in basic science, but also to a large number of unused patents. For example, an investigation of the FCC in 1934 reported that Bell owned or controlled 9,255 patents but actively used only 4,225 covered inventions (Antitrust Subcommittee, 1958, p.3842).

engineering workforce in the U.S. at the time.[6]

1.2.2 The Antitrust Lawsuit

On January 14, 1949 the United States Government filed an antitrust lawsuit with the aim to split AT&T from Western Electric.[7] The complaint charged that Western Electric and AT&T had been engaged in the monopolization of the manufacture, distribution and sale of telecommunications equipment in violation of the Sherman Antitrust Act of 1890 (Antitrust Subcommittee, 1959, p.46). According to the complaint, Bell was closing the market to all other buyers and sellers of telecommunications equipment by exclusionary practices including exclusive contracts and the refusal to license patents.[8]

To correct this, the government sought three main remedies. First, Western Electric was to be separated from AT&T, split into three competing companies, and to transfer all of its shares of the research subsidiary Bell Laboratories to AT&T. Second, AT&T was to buy telephone equipment only under competitive bidding and all exclusive contracts between AT&T and Western were to be prohibited. Third, the Bell System was to be forced to license all its patents for reasonable and non-discriminatory royalties (Antitrust Subcommittee, 1959, p.33).[9] Yet, none of this would happen.

The case ended with a consent decree on January 24, 1956, containing two remedies: First, the Bell System had to license all its patents issued prior to the decree royalty free to any applicant, with the exception of RCA, General Electric and Westinghouse who

---

[6]According to the National Science Foundation, the number of workers in S&E occupations was 182'000 in the U.S. in 1950. Source: https://www.nsf.gov/statistics/seind12/c3/c3h.htm - last accessed August 30, 2016.

[7]This account of facts follows largely the final report to the Antitrust Subcommittee of the House on the Bell Consent Decree Program (Antitrust Subcommittee, 1959).

[8]For example, Bell allegedly forced competitors "engaged in the rendition of telephone service to acquire AT&T patent license under threat of (...) patent infringement suits," or refused "to issue patent licenses except on condition" to be able to control the telephone manufacturer or by "refusing to authorize the manufacture (...) of telephones (...) under patents controlled by (...) the Bell System" or by "refusing to make available to the telegraphy industry the basic patents on the vacuum tube" that are essential for telegraphy to compete with telephone or by refusing to purchase equipment "under patents which are not controlled by Western or AT&T, which are known to be superior" (Antitrust Subcommittee, 1958, p.3838).

[9]There were two minor remedies: First, AT&T was not to be allowed to direct the Bell operating companies which equipment to purchase and second, all contracts that eliminated or restrained competition were to be ceased.

already had cross licensing agreements with Bell (the so called B-2 agreements). All subsequently published patents had to be licensed for reasonable royalties. As a consequence of the consent decree, 7'820 patents in 266 USPC technology classes and 35 technology subcategories or 1.3% of all unexpired U.S. patents became freely available. Second, the Bell System was barred from engaging in any business other than telecommunications.

The decree was hailed by antitrust officials as a "major victory", but already in 1957 the Antitrust Subcommittee of the Committee on the Judiciary House of Representatives started to investigate whether the decree of AT&T was in the public interest. The final report issued in 1959 pulled the decree to pieces: "the consent decree entered in the A.T. & T. case stands revealed as devoid of merit and ineffective as an instrument to accomplish the purposes of the antitrust laws. The decree not only permits continued control by A.T. & T. of Western, it fails to limit Western's role as the exclusive supplier of equipment to the Bell System, thereby continuing monopoly in the telephone equipment manufacturing industry."

The hearings of the Senate subcommittee uncovered a timeline of cozy back and forth negotiations and intense lobbying by the Department of Defense (DoD). The DoD intervened on behalf of Bell because it relied on the research of the Bell Labs. In World War II, the Bell Labs had been instrumental in inventing the superior radar systems of the Allies. They also engaged in around a thousand different projects, from tank radio communications to enciphering machines for scrambling secret messages (Gertner, 2012, p.59 ff.).[10] In the following years, Bell Labs continued to work for the DoD, for example by operating the Sandia National Laboratories, one of the main development facilities for nuclear weapons.

After the complaint was filed in January 1949, Bell sought and obtained a freeze of the antitrust lawsuit in early 1952 with support of the the DoD, on the grounds that Bell was necessary for the war effort in Korea. In January 1953, after Dwight D. Eisenhower took office, Bell began to lobby for the final dismissal of the case. The argument was that the Bell System was too important for national defense and thus should be kept intact. The government followed this argument and the Attorney General Herbert Brownell Jr.

---

[10]To highlight the engagement of Bell, we show in Figure A.2 in Appendix A the patenting activity of Bell in radar and cryptography during World War II.

asked Bell to submit concessions "with no real injury" that would be acceptable in order to settle (Antitrust Subcommittee, 1959, p.55)

In May 1954, AT&T presented and in June 1954 submitted to the Department of Justice a checklist of concessions that would be an acceptable basis for a consent decree. The only suggested major remedy was the compulsory licensing of all Bell patents for reasonable royalties. To support its position, Charles Erwin Wilson, the Secretary of Defense, wrote Herbert Brownell Jr., the Attorney General, a memorandum to the effect that the severance of Western Electric from Bell would be "contrary to the vital interests of our nation" (Antitrust Subcommittee, 1959, p. 56). In December 1955, the Department of Justice communicated with AT&T that it was ready to consider a decree of the "general character suggested [by A. T. & T.] in its memorandum (...) dated June 4, 1954" (Antitrust Subcommittee, 1959, p.92). Bell agreed.

1.2.3 Advantages of the Bell Case for the Empirical Set-up

The Bell case has two characteristics that make it ideally suited to measure the innovation effects of compulsory licensing as an antitrust remedy.

First, the consent decree did not impose any structural remedies for the telecommunications market. This allows us to isolate the innovation effect of compulsory licensing without any confounding changes in market structure. The reason why the Department of Justice did not impose any structural remedies is unclear. The final conclusion of the Antitrust Subcommittee blamed the lack of intent of the Attorney General to pursue Bell and the intense lobbying of the Department of Defense for the fact that no structural remedies were imposed (Antitrust Subcommittee, 1959, p.292). In contrast, the presiding judge Stanley N. Barnes stated that in his opinion it was enough to confine Bell to the regulated telecommunications market in order to prevent excessive prices and to end the exclusion of other suppliers (Antitrust Subcommittee, 1959, p.317).

Second, due to Bell Labs' commitment to basic science and its role in the war effort, Bell held a large number of patents unrelated to telecommunications, in industries in which it was not an active market participant. This gives us the opportunity to measure how

the innovation effect of compulsory licensing depends on the market structure. In the telecommunications industry, Bell was vertically integrated. Hence Bell was not only a dominant player in the production of the technology used for telephone equipment, but it also controlled the production of telephone equipment (Western Electric), as well as the product market for telephone equipment through its operating companies. In all other industries, Bell was a supplier of technology, but was not active in production. Even more, the consent decree explicitly banned Bell from ever entering into these businesses which meant that it effectively preserved the market structure inside and outside of the telecommunications industry.

To visualize the broad patent portfolio of Bell we use the data of Kerr (2008) to assign the most likely 4-digit SIC industry group to each USPC class (Figure 1.2). Around 42% of all Bell's patents have their most likely application in Bell's core business of producing telephones and telegraphs (SIC 3661). The remainder is spread across a large number of fields with an emphasis on electronics and industrial commercial machinery and computer equipment.[11]

1.3   Data and Empirical Strategy

For our estimation, we use comprehensive patent data for the U.S. from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. In this data, we identify all compulsorily licensed patents of the Bell System with a list of patent numbers published in the "Hearings before the Antitrust Subcommittee" of the U.S. Congress on the consent decree of Bell in May 1958 (Antitrust Subcommittee, 1958).[12]

In an ideal world, we would compare the number of realized follow-on innovations building on Bell patents with and without the consent decree. The problem is, however, that this is not possible: First, a census of follow-on innovations does not exist and second, we

---

[11]In Figure A.1 in Appendix A we show the compulsorily licensed patents split by technology subcategories following Hall et al. (2001). Only 31% of all Bell patents are in the field of telecommunications and the remaining patents are spread over 34 other subcategories.

[12]The list is the complete list of all patents owned by the Bell System in January 1956. It also includes patents of Typesetter Corp. which were explicitly excluded from compulsory licensing in Section X of the consent decree. We assume that these patents are unaffected.

Figure 1.2: Compulsorily Licensed Patents by Industry
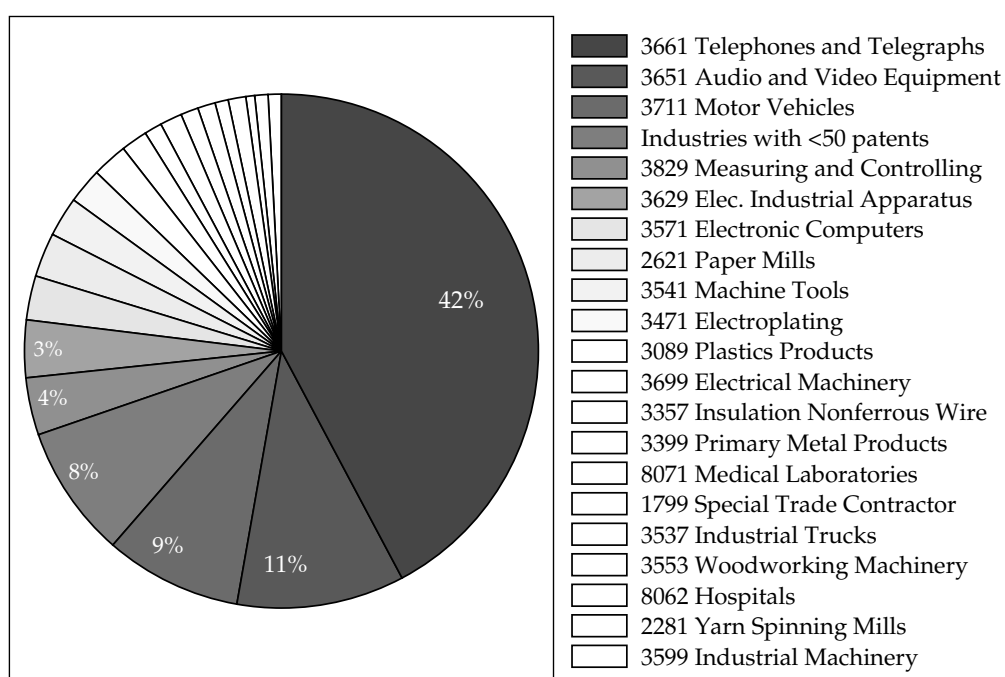
Notes: The pie chart shows the distribution of compulsorily licensed patents by most likely industry. We assign patents to the most likely 4-digit SIC industry using the data of Kerr (2008). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

can only observe the state of the world in which the compulsory licensing of Bell patents happened but not the counterfactual situation without the consent decree.

To measure follow-on innovations we use patent citations. Bell patents could be freely licensed after the consent decree, but patents that built on licensed Bell patents still had to cite them. Thus, we can use patent citations as a measure for follow-on innovations even though patents had lost their power to exclude competitors (Williams, 2015). The advantage of this measure is that, in contrast to most alternative measures such as new products or R&D spending, citations are consistently available from 1947 onward.[13] Citations have the additional advantage that they have a high frequency which allows a precise measurement of effects. The caveat is that some citations might have been added by the patent examiner, which adds noise to the measure (Alcacer and Gittelman, 2006; Alcacer et al., 2009).

To construct a counterfactual for the compulsorily licensed Bell patents, as control group we use all other patents that are published in the same year, that have the same total number of citations as the Bell patents in the five years prior to 1949, and that are in the same USPC technology class. By conditioning on the publication year and prior citations we control for the fact that, on average, young and high quality patents are cited more often. By conditioning on the same technology class we control for the number of companies that are active in the same field (i.e., for the number of potential follow-on inventors) and for technology-specific citation trends.

We can interpret our results causally under the assumption that in the absence of the consent decree, the Bell patents would have received the same number of citations as the control patents did (parallel trend assumption). More specifically, the identifying assumption is that conditioning on the control variables removes any systematic difference in follow-on citations between Bell and the control patents that is not due to compulsory licensing.

One potential concern about this identification strategy might be that the antitrust au-

---

[13]In 1947 the USPTO started to publish citations of prior art on the front page of the patent (Alcacer et al., 2009).The first patent to include prior art was issued on February 4, 1947. Yet, inventions were evaluated against the prior art already since the passage of the Patent Act of 1836. Prior to 1947, however, the prior art was available only from the "file history" of the issued patent, which is not contained in Patstat.

thorities chose to compulsorily license Bell patents for a reason related to the potential of follow-on research of these patents. According to the complaint and historical records, compulsory licensing was imposed because Bell used patents to block competitors in the field of telecommunications equipment. Therefore, if blocking patents are also patents that in the absence of compulsory licensing would have experienced particularly strong follow-on innovation, we might overestimate the effect of the consent decree.

Yet, this does not appear to be likely. In the absence of compulsory licensing, Bell's telecommunication patents would have continued to block competitors because the consent decree did not contain any other remedies aimed at restoring competition. Consequently, it seems fair to assume that blocking patents would have continued to receive the same number of citations as the control patents that have the same number of citations in the five years prior to 1949.

Furthermore, this concern obviously does not apply to the 58% of patents Bell held outside the field of telecommunications. These patents were included in the compulsory licensing regime of the consent decree not because they were blocking, but purely due to their association with the Bell System. Hence, there is no reason to expect any confounding effects.

To strengthen the point that the parallel trend assumption is plausible, we show in Section 1.4.1 that the number of citations of Bell and control patents was the same before the terms of the consent decree became known. In Section 1.4.3 we also show that companies that did not benefit from compulsory licensing did not start to cite Bell patents more after the consent decree. Thus, the control patents are a plausible counterfactual for patents both inside and outside of telecommunications.

Another concern might be that Bell's patenting strategy may have changed after the complaint became known. This is why we focus on patents *published* by 1949, the year the lawsuit against Bell started. The consent decree stated that only patents published before 1956 were to be compulsorily licensed. As a consequence of this cut-off date, more than 98% of the patents affected by the consent decree were filed before 1953, and more than 82% earlier than 1949. This implies that the characteristics of the majority of the affected patents were fixed before the Department of Justice filed its initial complaint.

To be on the safe side, we only use patents granted before 1949, but the results do not change when we use all patents affected by the consent decree.

Out of the 7'820 Bell patents affected by the Consent decree, 4'731 patents were published before 1949. For 4'533 of these patents (i.e., for 95.8%) we find in total 70'180 control patents that fulfill the criteria specified above. In our empirical analysis, we use the weights of Iacus et al. (2009) to account for the potentially different number of control patents per Bell patent.[14]

Table 1.1 shows summary statistics. In column (1) we report the summary statistics for all patents published between 1939 to 1956. In column (2) we report the summary statistics of all Bell patents that were published between 1939 and 1956 and hence were affected by the compulsory licensing rule. Patents published before 1939 had lost their patent protection by 1956 and were therefore not affected by the consent decree. In column (3) we report the summary statistics of the Bell patents published between 1939 and 1948. These are the patents that we use in our baseline regression.[15] They are affected by the consent decree but published before the lawsuit started and hence unaffected by a potential patenting policy change the lawsuit may have triggered.

The summary statistics of Bell patents differ from those of non-Bell patents. The average non-Bell patent in our data set receives 3.3 citations per patent and 6.1% of these citations are self-citations.[16] Bell System patents published in the same time period on average receive 5.2 citations and 13.4% of these citations are self-citations.[17] The numbers for the subsample of Bell patents published until 1949 are very similar. They receive on average 4.9 citations of which around 14.2% are self-citations.

---

[14]Iacus et al. (2009) proposes to use a weight of 1 for the treatment variable and a weight of $N_{Treatment,Strata}/N_{Control,Strata} \cdot N_{Control}/N_{Treatment}$ where $N_{Control}$ is the number of control patents in the sample, $N_{Control,Strata}$ is the number of control patents in a strata defined by the publication year, the USPC primary class and the number of citations up to 1949. $N_{Treatment}$ and $N_{Treatment,Strata}$ are defined analogously. Using these weights we arrive at an estimate for the average treatment effect on the treated.

[15]To make the statistics comparable for affected and not affected patents, we only consider technology classes in which Bell is active.

[16]In the main part of this chapter we only use citations by U.S. patents. In Appendix A, we run one regression with citations of patents filed in foreign jurisdictions.

[17]Except when explicitly mentioned in the text we correct for self-citations in all our regressions because we are mainly interested to which extent other companies built on Bell Labs patents.

Table 1.1: Summary Statistics

	(1) None-Bell System	(2) Bell System Affected	(3) Bell System Baseline Sample
	mean	mean	mean
Filing Year	1944.5	1943.6	1940.6
Publication Year	1947.6	1946.5	1943.1
# Years in patent protection after 1956	8.6	7.5	4.1
Total cites	3.3	5.2	4.9
Citations by other companies	3.1	4.5	4.3
Self Citations	0.2	0.7	0.7
Citations by other companies prior to 1949	0.3	0.9	1.4
Observations	293578	7820	4731

Notes: The table reports the average filing and publication year, the *average* number of years until patent expiration and citation statistics for patents published between 1939 and 1956. Column (1) includes all patents of non-Bell System companies in technologies where a Bell System company published at least one patent. Column (2) includes all Bell patents published between 1939 and 1956. Column (3) includes all Bell patents published between 1939 and 1949, the baseline sample of most of our regressions. A citation is identified as a self-cite if the applicant of the cited and citing patent is the same or if both patents belong to the Bell System. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

1.4 Results: Compulsory Licensing Increased Follow-on Innovation

Prior to the consent decree, Bell licensed its patents to other companies at royalty rates of 1% - 6% of the net sales price. Lower rates applied if a cross-license was agreed upon (Antitrust Subcommittee, 1958, p. 2685). The consent decree lowered these rates to zero and made licensing available without having to enter into a bargaining process with Bell. In this section we estimate whether and if so by how much this compulsory licensing increased follow-on innovations.

1.4.1 Timing: The Consent Decree Increased Citations of Other Companies Starting in 1955

In this section, we estimate the impact of the compulsory licensing on citations looking at the time period 1949-1970. We employ the following difference-in-differences specification:

Figure 1.3: Effect of Compulsory Licensing on Subsequent Citations
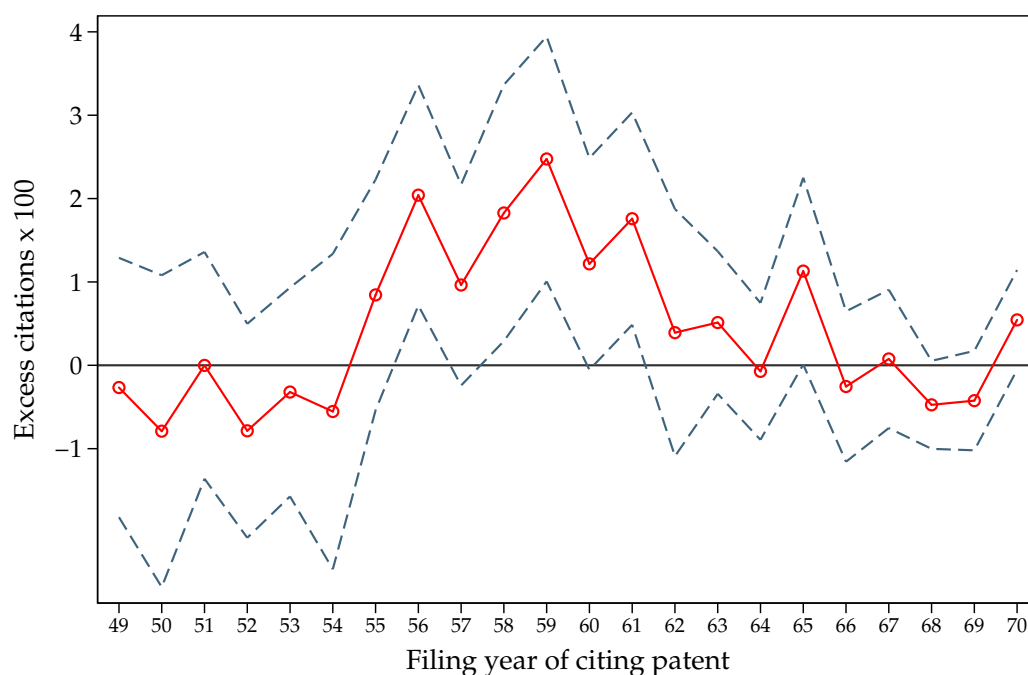
Notes: This graph shows the estimated number of yearly excess citations of patents affected by the consent decree ("Bell patents") relative to patents with the same publication year, in the same three-digit U. S. Patent Classification (USPC) primary class and with the same number of citations up to 1949. To arrive at these estimates we regress the number of citations in each year on an indicator variable that is equal to one if the patent under consideration is affected by the consent decree, and year fixed effects (Equation 1). We correct for self-citations. The dashed line represents the 90% confidence bands for the estimated coefficient. The sample under consideration contains 4'533 Bell patents and 70'180 control patents. We cannot match 198 Bell patents to control patents. To adjust for the different number of control patents per treatment patent in each stratum, we use the weights suggested by Iacus et al. (2009). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

$$\#Citations_{i,t} = \alpha + \beta_t \cdot Bell_i + YearFE_t + \varepsilon_{i,t} \qquad (1.1)$$

where $\#Citations_{i,t}$ is the number of follow-on citations of other companies to patent i in year t. $Bell_i$ indicates whether the patent i is owned by the Bell System and is therefore treated. We also include fixed effects for each year ($YearFE_t$).

Figure 1.3 shows per year the estimated number of excess citations of Bell patents that were granted before 1949 relative to control patents, β_t in equation 1.1. From 1949 to 1954, the average number of citations of treatment and control patents tracks each other very closely, speaking in favor of parallel trends in citations to Bell patents and to the control patents. In 1955, the average number of citations of other companies to Bell patents starts to increase and it converges again in 1960; 1960 is the average expiration date of the Bell patents in our sample.[18] The yearly coefficients from 1955 to 1960 are mostly significantly different from zero at the 10 % level.[19]

The increase in citations depicted in Figure 1.3 does not start in 1956, the year of the consent decree, but in 1955. This is plausible because on May 28, 1954, Bell already suggested a consent decree including the compulsory licensing of Bell System patents as described in Section 1.2. Thus, both the Bell Laboratories and companies building on Bell's patents could have known that compulsory licensing was pending as early as May 1954 (Antitrust Subcommittee, 1959).[20]

This timeline is supported by the cumulative abnormal stock returns for AT&T stocks shown in Figure 1.4. Up to the election of Dwight Eisenhower, cumulative abnormal returns were centered around zero. At the beginning of 1954, cumulative abnormal returns strongly increased to around 11%. The large uptick in March 1954 is exactly synchronized with the date of a memorandum summarizing a meeting of the Attorney General and Bell management about how to resolve the Bell case (Antitrust Subcommittee, 1958, p. 1956). Shortly thereafter, in May 1954, Bell proposed compulsory licensing as an acceptable remedy to settle the lawsuit. There is no more persistent positive or negative change in

---

[18]From 1861 to 1994, the term of the patent was 17 years from issuance.

[19]In the appendix to this chapter we graphically compare the average yearly number of citations to Bell and to control patents and find the same results.

[20]The first media mentioning of the consent decree against Bell was on May 13, 1955 in the New York Times. Public officials confirmed that top level negotiations are ongoing "looking towards a settlement of the AT&T case."

Figure 1.4: Cumulative Abnormal Stock Returns of AT&T
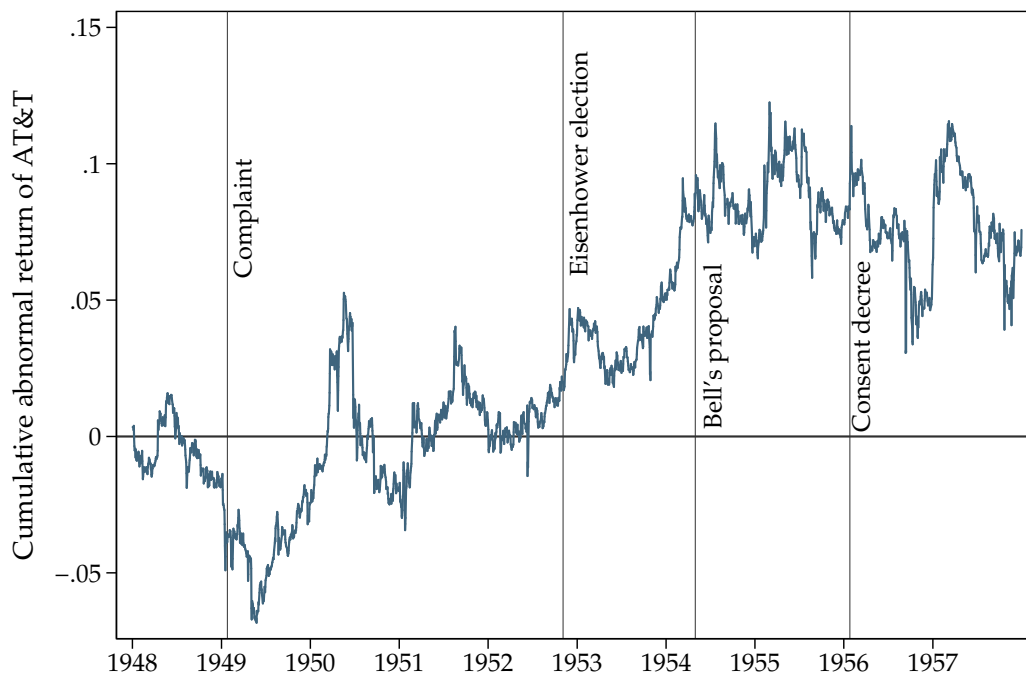
Notes: This figure shows the cumulative abnormal stock return of AT&T compared to other companies in the Dow Jones index, beginning in January 1948. The events marked in the graph are the beginning of the antitrust lawsuit on January 14, 1949, the presidential election on November 4, 1952, Bell's proposal of compulsory licensing on June 4, 1954, and the consent decree on January 25, 1956. The data are from the Center for Research in Security Prices (CRSP).

the cumulative abnormal return until 1959. In particular, the consent decree itself in 1956 did not seem to have had any more informational value.

We can also infer from Bell's behavior that as early as the first half of 1955, compulsory licensing was expected. According to the consent decree, all patents had to be licensed for free if they were published before January 24, 1956. If they were published after this cut-off date, they were licensed on a reasonable and non-discriminatory basis. So starting from the date when Bell became aware of the clause it had an incentive to delay the publication of its patents beyond the cut-off date.

According to the data, Bell indeed started to delay its patents at the patent office beginning in the first half of 1955. To pin down the date, we compare the propensity of a Bell patent to be published with the propensity that control patents are published for a given filing year. In Figure 1.5, we show these hazard rates of publishing in a particular

year for the filing years 1949 and 1953.[21] For the filing year 1949, the publishing rates per year are very similar for Bell patents and patents from other companies. If at all, Bell patents were published a bit earlier. For the filing year 1953, this picture is reversed: Starting in the first half of 1955, Bell patents had a significantly lower probability of being published. This is consistent with Bell trying to delay the publications of its patents and having credible information about the general outline of the consent decree in the first half of 1955 at the latest.

1.4.2 Magnitude: The Consent Decree Increased Citations to Bell Patents by 17%

We next present our baseline regression. To quantify the size of the effects of the consent decree, we estimate the average yearly effect of the consent decree on citations of other companies for the time period 1949-1960. We employ the following difference-in-differences model:

$$\#Citations_{i,t} = \beta_1 \cdot Bell_i + \beta_2 \cdot I[1955 - 1960] + \beta_3 \cdot Bell_i \cdot I[1955 - 1960] + \varepsilon_{i,t} \quad (1.2)$$

where $I[1955 - 1960]$ is an indicator variable for the treatment period. We define the treatment period as from 1955 to 1960 based on the yearly coefficients in Figure 1.3.

The results are reported in Table 1.2 column (1).[22] In the treatment period, the consent decree resulted in 0.020 additional citations. This implies that, on average, the consent decree increased citations to Bell patents by other companies by 17% from 1955 to 1960.[23] Considering only the 4'731 patents published before 1949, this implies a total increase of 568 citations. If we assume homogeneous effects for all 7'820 patents published up to 1956, the total number of excess citations is 938. The effect is also positive and statistically significant if we include all patents up to 1956, the year of the consent decree (column 2).

---

[21]Hazard rates for all other years are available from the authors upon request.

[22]Note that patents receive fewer citations post treatment because older patents in general receive fewer citations than younger patents. See Figure A.3 in Appendix A.1.

[23]To determine the percentage increase, we first calculate the number of citations Bell patents would have received in the absence of the treatment (counterfactual), using the coefficients in Table 1.2, column (1). The counterfactual is 0.115 (= 0.183 - 0.004 - 0.064). We then divide the treatment effect, 0.02, by the counterfactual (0.02/0.115 = 0.174).

Figure 1.5: Hazard Rates for Publication of Patents by Filing Year

(a) Filing year 1949
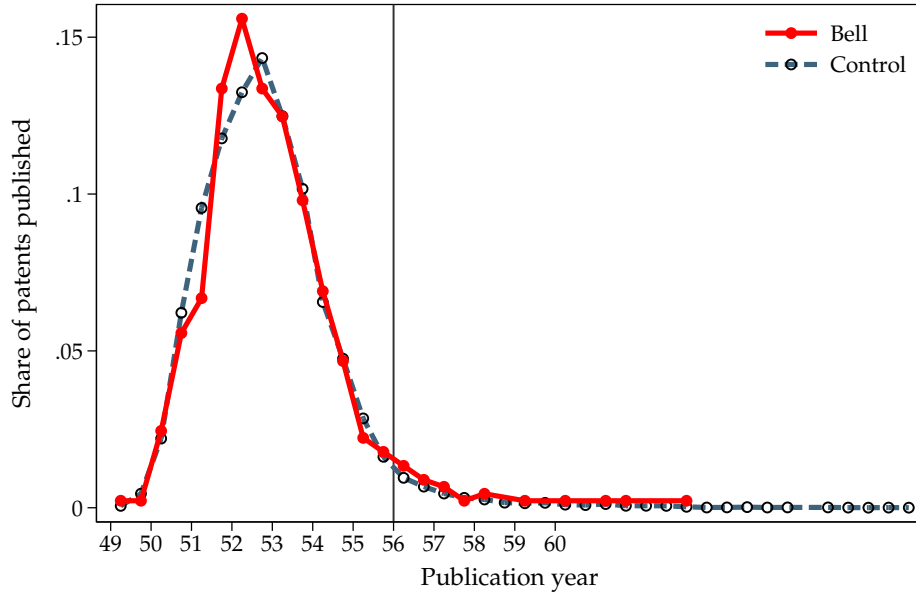
(b) Filing year 1953
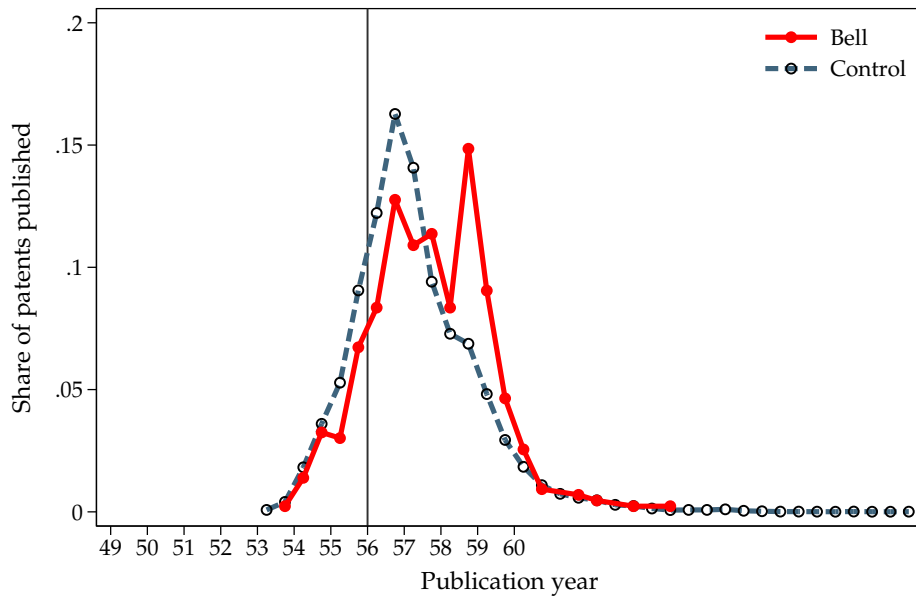
Notes: These figures show the hazard rates for publication of patents that were filed by Bell (solid line) and others (dotted line). Subfigure (a) shows hazard rates for patent applications filed in 1949, subfigure (b) for applications filed in 1953. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Back-of-the-envelope calculations suggest that the additional patents for other companies directly induced by the consent decree had a total value of up to $5.7 billion. To calculate this number we use estimates for the average dollar value derived from Kogan et al. (2017) to weigh each citing patent.[24] According to these estimates, each compulsorily licensed patent created an additional value of $121'000 annually in the treatment period (column 3). Assuming homogeneous effects for all 7'820 patents in the treatment group, the consent decree led to around $5.7 billion in economic value over six years, between 1954 and 1960. These calculations represent an upper bound because they assume that without the additional citations induced by the consent decree, the patent would not have been invented (i.e., that the compulsorily licensed patent was strictly necessary for the citing invention).

The effect is measurable across the quality distribution of patents. We split all patents by the number of citations a patent received in the first five years after publication and present results in columns (4) and (5) of Table 1.2. We define a high-quality patent as a patent with at least one citation before 1949 and a low-quality patent as a patent with no citations. The effect is stronger for high quality patents, but the effect is also statistically significantly different from zero for low quality patents. The effect is also not exclusively driven by the computer industry, which was just about to start in 1956. In column (6), we report results when dropping all 491 Bell patents classified in the technology subcategories "Computer Hardware and Software", "Computer Peripherals" and "Information Storage" or "Others" (Hall et al., 2001) and find a similar effect. The effect is also not driven by the concurrent consent decrees of IBM in 1956 or RCA in 1958. IBM and RCA were defendants in an antitrust case with compulsory licensing as the outcome. We drop all citations from patents that also cite either the patents of RCA or the patents of IBM and report the results in column (7).

---

[24]Kogan et al. (2017) measure the value of a patent using abnormal stock returns around the publishing date of the patent. We use this data to calculate the average dollar value for a patent in each technology class and publication year.

Table 1.2: The Effect of Compulsory Licensing on Subsequent Citations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Citations by			Bell	all
				other companies					
	Baseline	Up to 1956	Dollar weighted	Low quality	High quality	w/o Computer	w/o IBM & RCA	Self-Cites	Total-Cites
Bell	-0.4	-0.6***	0.1	-0.5	-0.3	-0.5	-0.6	1.4***	0.5
	(0.5)	(0.2)	(2.9)	(0.4)	(0.9)	(0.5)	(0.5)	(0.3)	(0.7)
I(55-60)	-6.4***	-3.3***	3.7	-0.5**	-12.4***	-6.9***	-6.2***	-1.0***	-6.8***
	(0.6)	(0.8)	(2.3)	(0.2)	(0.9)	(0.7)	(0.6)	(0.1)	(0.7)
Bell x I(55-60)	2.0***	1.9***	12.1***	1.0**	3.1***	2.2***	2.0***	-0.6**	1.6*
	(0.6)	(0.6)	(4.0)	(0.4)	(1.0)	(0.6)	(0.6)	(0.3)	(0.8)
Constant	18.3***	19.9***	83.2***	8.4***	28.2***	18.7***	17.6***	1.5***	19.0***
	(1.2)	(1.6)	(3.7)	(0.4)	(1.4)	(1.4)	(1.1)	(0.1)	(1.2)
# treated	4533	7111	4533	2279	2254	4042	4533	4444	4731
Clusters	225	253	225	194	179	160	225	223	223
Obs.	896556	1121648	896556	580356	316200	700500	896556	854592	828876

Notes: This table shows the results from a difference-in-differences estimation with 1949-1954 as pre-treatment period and 1955-1960 as treatment period. The estimation equation is:

$$\#Citations_{i,t} = \beta_1 \cdot Bell_i + \beta_2 \cdot I[1955-1960] + \beta_3 \cdot Bell_i \cdot I[1955-1960] + \varepsilon_{i,t}$$

where $I[1955-1960]$ is an indicator variable for the treatment period from 1955 to 1960. *Bell* is an indicator variable equal to one if a patent is published by a Bell System company before 1949. As control patents we use all patents that were published in the U.S., matched by publication year, primary USPC technology class, and the number of citations up to 1949. To adjust for the different number of control patents per treatment patent in each stratum, we use the weights suggested by Iacus et al. (2009). As dependent variable, we use all citations by companies other than the filing company in columns (1) through (7). In the second column, we extend our sample of affected patents to 1956, and in the third column we use the sample up to 1949 and weight each citation by the average dollar value of a patent in the same publication year and technology class derived from the values provided by Kogan et al. (2017). In columns (4) and (5), we split the sample by their citations prior to 1955 to measure quality of patents. A patent is classified as "high quality" if it has at least one citation prior to 1955 and it is classified as "low quality" otherwise. In column (6) we exclude patents that are classified in technology subcategories related to the computer, and in column (7), we exclude all citations by patents of IBM and RCA and all patents that cite IBM and RCA patents. IBM had a consent decree with a compulsory licensing of patents in 1956 as well and RCA had a consent decree in 1958. In column (8), we match patents on publication year, technology class and self-citations as outcomes. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. All coefficients are multiplied by 100 for better readability. Standard errors are clustered on the primary three-digit USPC technology class level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

1.4.3 Robustness Check: No Increase in Citations by Untreated Companies

One concern for the estimation is that the effect of compulsory licensing on subsequent citations might be driven by a shock that increased follow-on innovation to Bell patents and was correlated with the consent decree. For example, the antitrust prosecutors might have chosen to press for compulsory licensing because they expected that there would be many follow-on innovations based on the high quality of Bell's patents.

To see whether this might have been the case we analyze the citation patterns of unaffected companies to Bell patents and to the control patents. The 1956 consent decree singled out three companies that were explicitly excluded from the free compulsory licensing of Bell patents: the General Electric Company, Radio Corporation of America, and Westinghouse Electric Corporation. The reason was that these companies already had a general cross-licensing agreement, the "B-2 agreements" dated July 1, 1932. A fourth company, the International Telephone and Telegraph Company (ITT), was also not affected by the decree as it had a patent pool with Bell.

We repeat our baseline analysis but use only the citations of the B-2 companies (including ITT) as the dependent variable and report the results in Figure 1.6 and column (2) of Table A.1 in Appendix A.1. We do not find any effect. This suggests that the consent decree did not change the citation behavior of excluded companies and the measured effects are not due to a common technology shock. As these companies in total make up 12% of all citations to Bell patents, this null effect is not due to a lack of measurability.[25]

A second concern might be that due to the free availability of Bell technology, companies substituted away from other, potentially more expensive technologies. In Appendix A.1 we show the results of additional auxiliary analyses suggesting that the effects are not driven by citation substitution.

---

[25]We repeat our analysis also for foreign companies, which could also use Bell patents for free but which did not receive technical assistance, and report the results in Table A.1, column (3) in Appendix A.1. Similarly, we repeat our analysis for companies that already had a licensing agreement in place and compare them with companies without a licensing agreement (Table A.1, columns 4 and 5, Appendix A.1). As expected, we find that the effects are smaller for firms that were less affected by the consent decree.

Figure 1.6: Effect of Compulsory Licensing on Subsequent Citations Among Companies Exempt from the Consent Decree
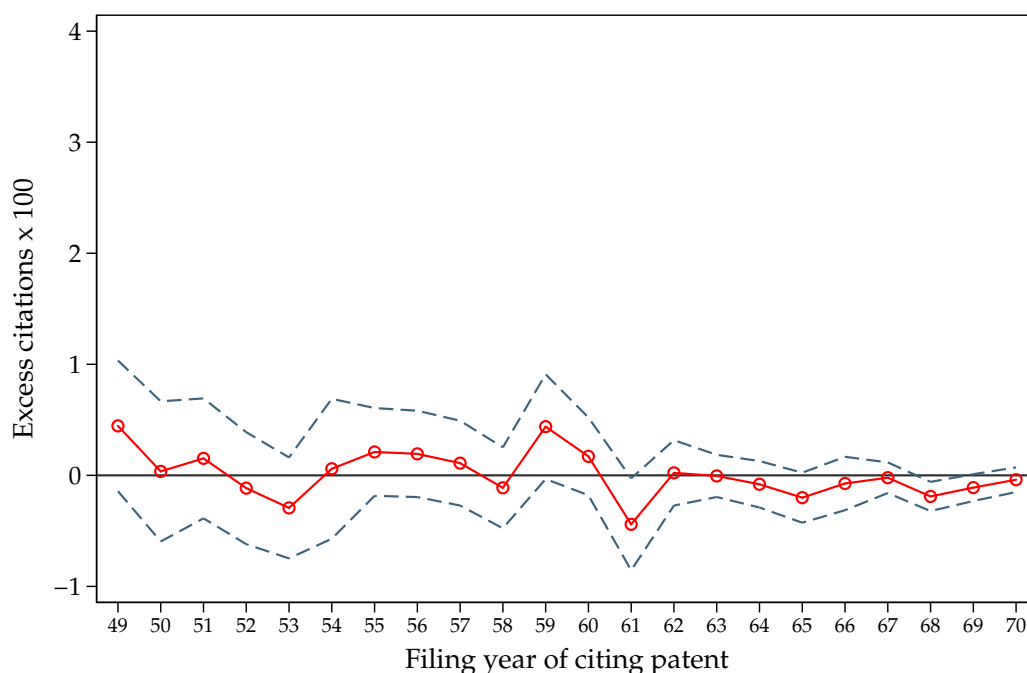
Notes: This graph shows the estimated number of yearly excess citations by General Electric Company, Radio Corporation of America and Westinghouse Electric Corporation, the three companies exempt from the consent decree, and by International Telephone and Telegraph Company, which already had a patent pool in place, of patents affected ("Bell patents") relative to patents with the same publication year, in the same three-digit USPC primary class and with the same number of citations up to 1949. To arrive at these estimates, we regress the number of citations by the unaffected companies in each year on an indicator variable equal to one if the patent under consideration is affected by the consent decree and year fixed effects. The dashed line represents the 90% confidence bands for the estimated coefficient. To adjust for the different number of control patents per treatment patent in each stratum, we use the weights suggested by Iacus et al. (2009). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Finally, in Appendix A.1 we vary the construction of control groups and show that our results are not driven by the particular choice of matching variables.

1.4.4 Robustness Check: The Decrease in Bell's Own Patenting is Lower than the Increase in Patenting by Other Companies

We next examine how Bell reacted to the consent decree. Bell might have reduced its innovation activities by more than other companies increased their innovation activities, such that the net-effect of the consent decree would be negative. To see whether this is the case, we measure whether Bell continued to produce follow-on innovations building on its own patents.[26] Results are reported in column (8) of Table 1.2. The number of self-citations shows a decrease of 0.006 self-citations in the years between 1955 and 1960. This decrease is statistically significant, but is not large enough to dominate the increase in citations by other companies. In column (9) we present the effect on total citations, i.e., citations by other companies and self-citations by Bell. We find that total citations increased by 0.016. This speaks in favor of a net increase in innovation due to the consent decree.

Bell's innovation output in terms of number of patents continued to grow in line with expectations in the years following the consent decree. To show this, we construct a synthetic Bell and compare it with the actual patent output of the Bell System. To construct a synthetic Bell, we first calculate the share of Bell's patents of all patents in each technology subcategory for the years 1946, 1947, and 1948. Then we assume that Bell's growth would have been in line with the growth of other companies that existed before 1949 in these technology subcategories so that Bell would have held its share in each subcategory constant for the following years. Results are presented in Figure 1.7a. It shows that Bell's patenting is on average smaller than the patenting of the synthetic control, but not by much.[27]

---

[26]Self-citations are a measure for how much a company develops its own patents further (Akcigit and Kerr, forthcoming; Galasso and Schankerman, forthcoming).

[27]In Figure A.5 in the Appendix A.1 we compare the patenting output of Bell with other control companies and find that Bell's patent growth is in line - but at the lower end - of similar companies.

Figure 1.7: Innovation and R&D in the Bell System After the Consent Decree

(a) Patenting Over Time: Bell system and Synthetic Bell
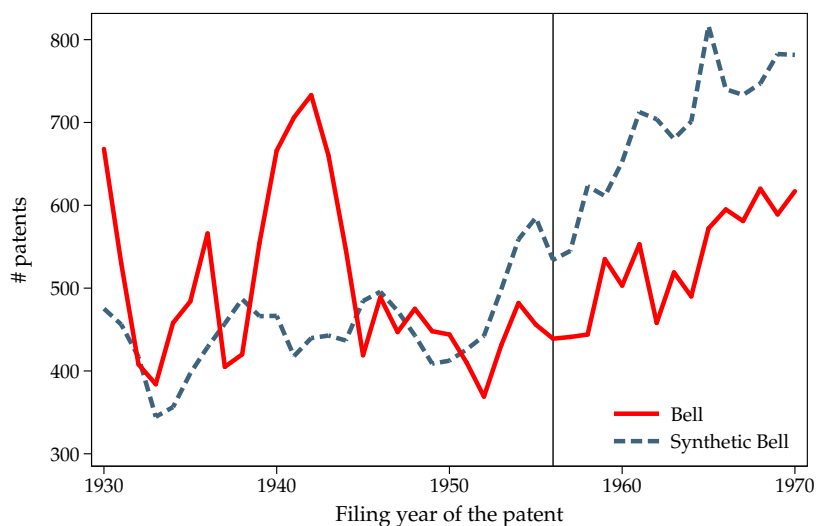
(b) R&D Spending by AT&T
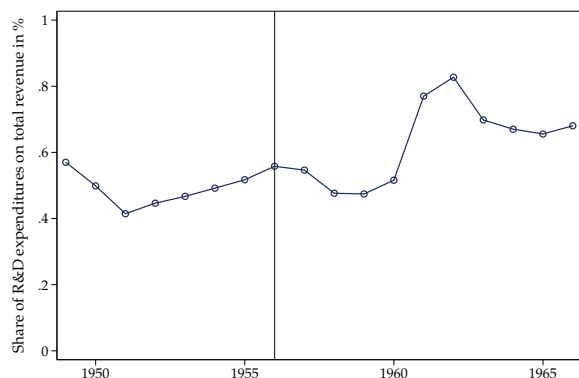
(c) Share of Communication Patents
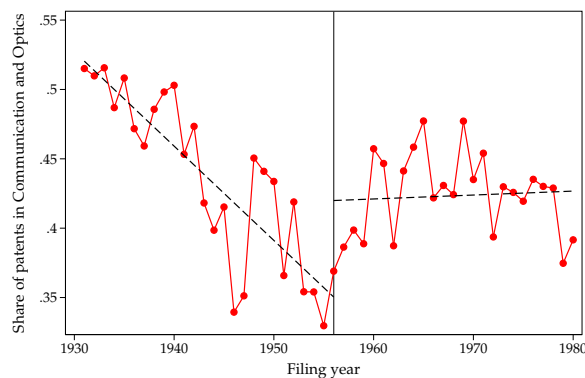
Notes: Subfigure (a) shows the total number of patents filed by the Bell System compared to a synthetic Bell. To construct the synthetic Bell, we calculate the share Bell's patents had in each 2-digit technology subcategory relative to all patents of companies that had at least one patent before 1949. We then assume that in the absence of the consent decree, Bell's patenting would have grown in each subcategory at the same pace as the patenting of all other companies. As a consequence, Bell's share in each technology subcategory is held constant. In a last step, we add the number of patents up to a yearly sum. Subfigure (b) shows the ratio of R&D expenditures relative to total R&D of American Telephone & Telegraph. The data are from the annual reports of AT&T. Subfigure (c) shows the share of patents related to communication relative to all patents filed by Bell. We define a patent as related to communication if the most likely application is in the production of telecommunications equipment (SIC 3661). In Appendix A, we show the change in direction using NBER subcategories. The patent data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Bell's continued investment in research was in line with the incentives the consent decree and the regulators provided. The consent decree did not significantly alter the profitability of new patents. The consent decree mandated that Bell could demand "reasonable" licensing fees for all patents published after January 1956. The reasonable royalty rates Bell charged were not much different compared to the pre-decree royalties (Antitrust Subcommittee, 1959, p.111). The only difference was that Bell had to give a license to any applicant.

Bell also had little incentive to reduce investment in R&D because the Bell System was subject to a rate of return regulation following the Communications Act of 1934. According to annual reports, AT&T had a stable ratio of R&D to operating revenue of 0.5% from 1949 to 1960 (Figure 1.7b).[28] For the entire Bell System, the share of R&D to total turnover stayed almost constant at 2%-3% from 1966 to 1982 (Noll, 1987). However, the absolute level of R&D effort increased as the Bell System grew. Operating revenues increased from $3.2 billion in 1950 to $5.3 billion in 1955, to $7.3 billion in 1960 and to $11 billion in 1965, while the staff at Bell Labs grew from 6'000 in 1950, to 10'000 in 1955, to 12'000 in 1960 and 15'000 in 1965 (Temin and Galambos, 1987).

But even if the consent decree offered no incentive for Bell to downsize, it offered incentives for Bell to redirect its research budget towards applications in the telecommunications field. Prior to the consent decree, Bell could expand to other businesses. Afterward, Bell's future was bound to common carrier telecommunications. The company correspondingly refocused its research program on its core business and increased its share of patents in fields related to the production of telecommunications equipment (Figure 1.7c).

These results are consistent with the study of Galasso and Schankerman (forthcoming) on patent invalidations. They show that large companies on average do not reduce follow-on innovations significantly if they lose a patent due to litigation. The only exception is if the large company loses a patent outside of its core-fields. Then it reduces innovation in the field of the patent under consideration and reacts by redirecting future innovation to

---

The only exception is the growth of General Electric which is much larger, highlighting the problem of constructing a counterfactual for a single company.

[28]We do not know whether the consolidated balance sheet also includes the Bell Laboratories and Western Electric. It seems that at least some parts of the Bell System are not consolidated in the annual reports of AT&T.

a different but related field.

1.4.5 Mechanism: Increase in Citations is Driven by Start-ups

We next examine which type of company increases innovation after the compulsory licensing and report the results in Table 1.3. We split citations by the type of the citing assignee. An assignee is either a company or an individual inventor; an assignee is defined as young and small if its first patent was filed less than 10 years before it cited the Bell patent and if it had less than 10 patents before 1949.[29] We first use the number of citations from young and small assignees as the dependent variable and report the results in column (2). We then use the citations of all other assignees that are not young and small and report the results in column (3). In column (4) we look explicitly at small and young assignees that are companies ("start-ups"), leaving out individual inventors.[30]

We find that the increase in follow-on innovation is predominantly driven by young and small companies entering the market and by individual inventors. Young and small assignees increase their citations after 1955 by an average of 0.014 citations (32%) while all others increase their citations by 0.006 (6%) on average. Around 70% of the overall increase comes from young and small assignees, but they are responsible for only one-third of all citations to Bell patents (columns 2 and 3 in Table 1.3).[31] Among the small and young assignees, start-ups experience a particularly strong increase: they account for 50% of the total increase in citations although they are responsible for only 18% of all citations (column 4).

---

[29]In Appendix A.1 we use different definitions for young and small companies and find that the effect is mainly driven by companies that file their first patent.

[30]We identify companies as all assignees that are never inventors. Our results are robust to defining companies as having Inc., Corp., Co. or similar abbreviations in their name.

[31]Young and small assignees are responsible for an increase of 0.014 citations (column 2). This is 70% of the total increase of 0.02 (column 1). It is also an increase of around 32% relative to what we would have expected without a consent decree. According to the estimates a Bell patent should have received 0.044 citations (0.068 is the constant, the Bell effect is -0.008, and the average decrease in citations in the post treatment period is -0.016) but did receive 0.058 citations (0.044 baseline effect + 0.014 treatment effect).

Table 1.3: The Effect of Compulsory Licensing on Subsequent Citations by Company Type and Field

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Base line	Age and Size		Y&S Comp	Former Bell?		Citations to Y&S by	
		Young & Small	Others		No	Yes	Others	Communication
Bell	-0.4	-0.8***	0.4	-0.5**	-1.2**	0.8***	-0.6***	-0.2
	(0.5)	(0.3)	(0.5)	(0.2)	(0.5)	(0.2)	(0.2)	(0.3)
I(55-60)	-6.4***	-1.6***	-2.7***	-0.5*	-5.9***	-0.5***	-1.1***	-0.6*
	(0.6)	(0.3)	(0.8)	(0.2)	(0.6)	(0.1)	(0.2)	(0.3)
Bell x I(55-60)	2.0***	1.4***	0.6	1.0***	2.5***	-0.5**	1.1***	0.3
	(0.6)	(0.3)	(0.6)	(0.3)	(0.5)	(0.2)	(0.3)	(0.2)
Constant	18.3***	6.8***	12.4***	3.7***	17.2***	1.1***	5.2***	1.7***
	(1.2)	(0.4)	(1.0)	(0.2)	(1.2)	(0.1)	(0.7)	(0.5)
# treated	4533	4533	4533	4533	4533	4533	4533	4533
Clusters	225	225	225	225	225	225	225	225
Obs.	896556	896556	896556	896556	896556	896556	896556	896556

Notes: This table shows the results from a difference-in-differences estimation with 1949-1954 as the pre-treatment period and 1955-1960 as the treatment period. The variable Bell is an indicator variable equal to one if a patent is secured by a Bell System company before 1949 and therefore a subject to the consent decree. As control patents we use all patents that were secured in the U.S., matched by publication year, primary USPC technology class, and the number of citations up to 1949. To adjust for the different number of control patents per treatment patent in each stratum, we use the weights suggested by Iacus et al. (2009). As dependent variable, we use all citations by companies other than the filing companies in column (1). We split these citations according to the age and size of the company. In column (2), we use only citations by young and small inventors, defined as having applied for their first patent no more than ten years ago and having less than ten patents overall. In column (3), we use only the citations of inventors that are neither young nor small and in column (4) of companies that are both young and small. In columns (5) and (6), we split the citations according to whether inventors ever patented for Bell or ever were co-authors with Bell inventors. In columns (7) and (8), we split citations coming from inside or outside of the communication field. We determine whether a citing patent is inside the communication field if the technology class has the most likely application in the production of telecommunications equipment (SIC 3661), using the data of Kerr (2008). The patent data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. All coefficients are multiplied by 100 for better readability. Standard errors are clustered on the primary three-digit USPC technology class level and *, **, *** denote statistical significance on 10%, 5% and 1% level, respectively.

These results suggest that patents act as a barrier of entry for start-ups and prevent their follow-on innovation. They provide support for the hypothesis that the consent decree reduced potential bargaining failures. Several prior studies suggest that small firms might not have large enough patent portfolios to resolve disputes or to strike cross-licensing agreements (Lanjouw and Schankerman, 2004; Galasso, 2012; Galasso and Schankerman, 2015). As cross-licensing was a priority in the licensing strategy of Bell prior to the consent decree, a small patent portfolio might have been a significant handicap for small inventors seeking a license from Bell (Antitrust Subcommittee, 1958, p. 2685).

One potential concern might be that the observed increase of citations by young and small companies was not driven by the consent decree itself but by other changes at Bell Laboratories. Historical accounts suggest that there was an exodus of important Bell researchers around the time of the consent decree. For example, in 1953 Gordon Teal, inventor of a method to improve transistor performance, joined the then small Texas Instruments Inc. Similarly, William Shockley, one of the inventors of the transistor, left Bell in 1956 to found Shockley Semiconductors Laboratory.

To show that this is not the case, we separately look at patent citations by people who were at some point associated with Bell, but later patented for a different company, including their co-inventors, and compare with citations by all remaining unrelated inventors. In our data, there are 4'477 former Bell employees with 28'569 patents. These people have in total 12'068 co-inventors who were never active at Bell and who filed 87'148 patents in total. The results are reported in columns (5) and (6) of Table 1.3. We find a positive effect on the citations of unrelated inventors and a negative effect on the citations of related inventors.[32] This pattern does not suggest that the increase in follow-on innovation was driven by former Bell employees. However, the results do suggest that Bell's inventors had preferential access to Bell technology prior to the consent decree and that there was a strong increase from unrelated inventors afterwards.

---

[32]The estimated yearly coefficients for excess citations of former Bell inventors and of unrelated inventors are available from the authors upon request.

1.5 Compulsory Licensing did not End Foreclosure in the Market for Telecommunications Equipment

The aim of the consent decree was to end foreclosure in the market for telecommunications equipment. According to the antitrust lawsuit, Bell was closing the market to all other buyers and sellers of telecommunications equipment by using exclusive contracts between Western Electric and the Bell operating companies and by refusing to license patents to competitors. In markets outside of the telecommunications industry, Bell was active only as a supplier of technology but was not an active market participant.

Market foreclosure is thought to have a negative effect on the innovation activities of the companies that are foreclosed (Baker, 2012; Wu, 2012). The argument is that foreclosed companies cannot earn profits by selling their improved products directly to consumers. The only option they have is to sell their innovations to other companies.[33] Thus, foreclosed companies have lower incentives for innovation than companies with access to a customer base.

In this section, we compare the innovation effects of compulsory licensing inside and outside of the telecommunications industry to infer whether market foreclosure is harmful for innovation and whether compulsory licensing is effective in ending it. If compulsory licensing increases innovation in the same way in all industries, then any difference between the two competitive settings must be due to market foreclosure in the telecommunications industry. If market foreclosure reduces innovation as argued above and if compulsory licensing was effective in ending it, we should see a stronger increase in follow-on innovations in the telecommunications industry than in other industries. In contrast, if compulsory licensing was ineffective in ending market foreclosure, we should find a smaller effect. If market foreclosure has no effect on innovation, we should find similar effects in all industries.

To compare the innovation effects within telecommunications and outside we first need to characterize each citing patent by its closeness to the market for telecommunications

---

[33]Such a market for ideas exists only in special circumstances (Gans et al., 2002; Gans and Stern, 2003; Gans et al., 2008).

Figure 1.8: Excess Citations by Patents with Varying Likelihood of Being Used in Production of Communication Equipment
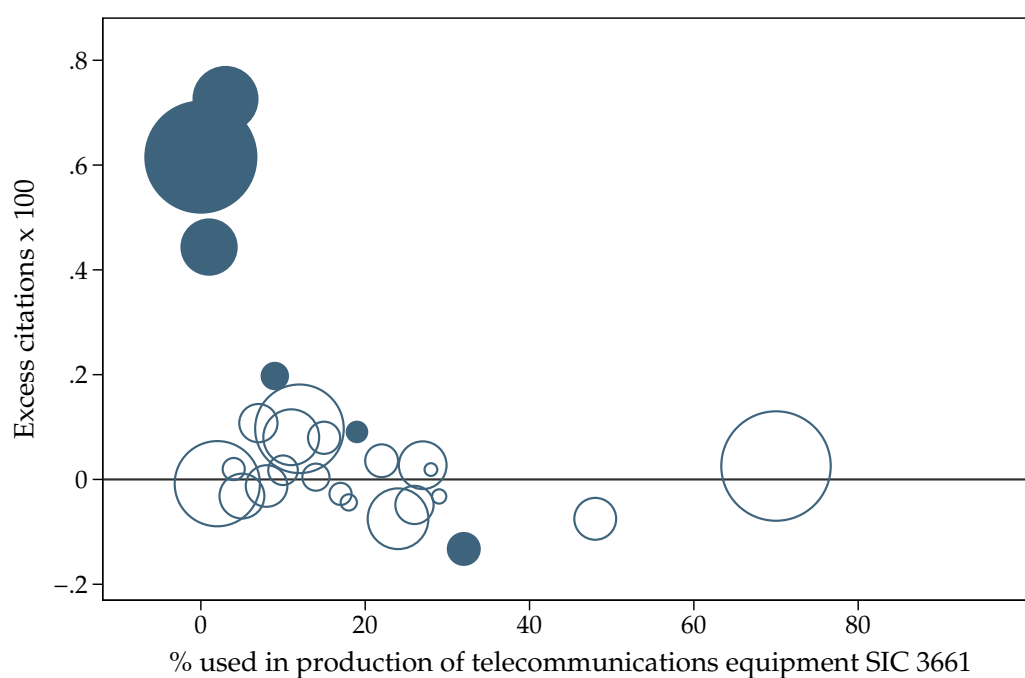
Notes: This figure shows results from a difference-in-differences estimation of the impact of the consent decree on follow-on patent citations with 1949-1954 as the pre-treatment period and 1955-1960 as the treatment period, controlling for year fixed effects. We estimate Equation 1.2 and report β_3 separately, using as dependent variables citations from patents with a different relevance for the production of telecommunication equipment (SIC 3661 - "Telephone and Telegraph Apparatus"). Relevance is measured by the likelihood that a patent is used in industry SIC 3661 using the data of Kerr (2008). The size of the circle signifies the number of Bell patents in a technology and a solid circle implies that the coefficient is significant at the 10% level. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

equipment. To do this, we use the concordance of Kerr (2008) that gives us the probability for each USPC technology class that a patent in this technology class is used in the production of telecommunications equipment (SIC 3661). We interpret this probability as a measure of closeness to telecommunications. We then assign this probability to each citing patent according to its technology class and sum up the citations for each level of likelihood to construct a different dependent variable for each level of closeness, 26 altogether. In a last step, we repeat our main regression for each level of closeness. We can thus estimate how much the consent decree increased citations in markets that are close to the production of telecommunications equipment and in markets unrelated to it.

In Figure 1.8 we show the average treatment effects estimated with our baseline model in equation 1.2 for different levels of closeness to the production of telecommunications equipment. We find a strong negative relation between the closeness to telecommunications and excess citations. Almost all excess citations come from patents that have nothing to do with telecommunications. We conclude that follow-on innovation in telecommunications was not influenced by compulsory licensing. Under the assumption that compulsory licensing affects innovation similarly in all industries this result supports the argument that market foreclosure has a chilling effect on innovation and indicates that compulsory licensing was ineffective in solving it.

Next, we use Kerr's data to assign each citing patent to the industry in which it is most likely used and repeat the baseline regression with citations from patents in different industries. The results are shown in Figure 1.9. Almost all additional citations are from patents with the most likely application outside of the industry "Telephones and Telegraphs" (SIC 3661). A large part of the effect is driven by unrelated industries such as "Measuring and Controlling," "Audio and Video Equipment" or "Motor Vehicles".[34] These results support the notion that market foreclosure is harmful for innovation and that compulsory licensing is ineffective as a remedy.[35]

---

[34]In the Appendix A.2 we repeat the analysis using NBER technology subcategories to classify the citing patent. The results are the same.

[35]Another explanation for our null result in the telecommunications market would be that there was a lack of innovation potential in the telecommunication sector after 1956. To rule out this hypothetical possibility we compare the development of patents in the telecommunications sector. Results are reported in Figure A.9 in Appendix A.2 They show that the number of citations to Bell's telecommunications patents had a similar trend as patents outside of telecommunications and that the number of Bell's newly filed telecommunications patents shows no signs of abating after the consent decree.

Figure 1.9: Excess Citations by Patents According to the Most Likely SIC Industry Classification
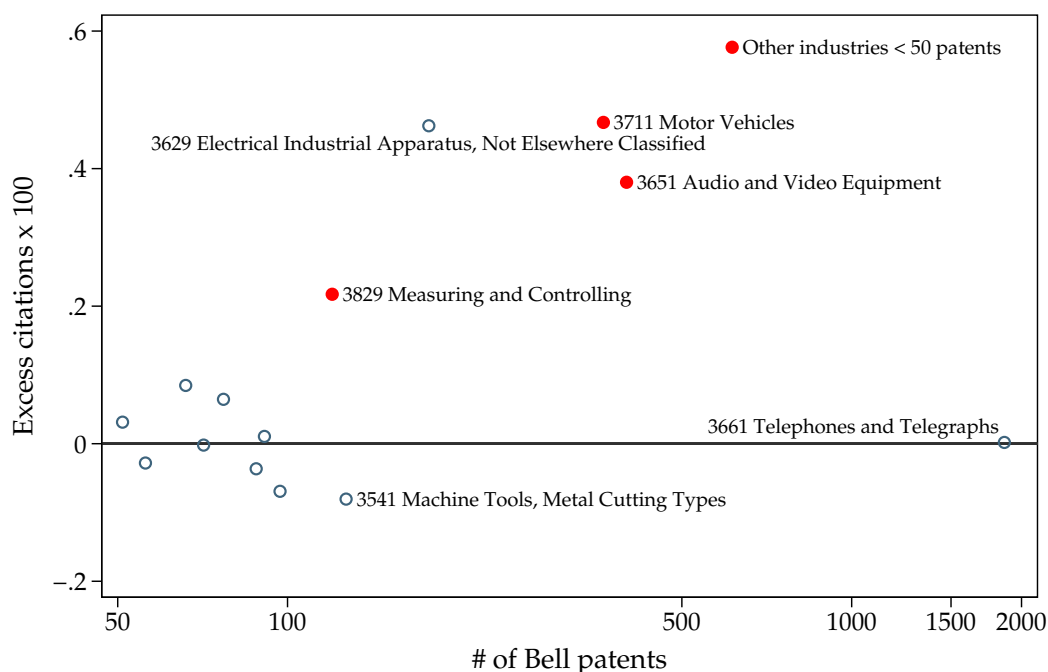
Notes: This figure shows results from a difference-in-differences estimation of the impact of the consent decree on follow-on innovation with 1949-1954 as the pre-treatment period and 1955-1960 as the treatment period, controlling for year fixed effects. As the dependent variable, we use all citations by companies other than the filing companies classified by the most likely SIC classification of the citing patent. As control patents, we use all patents that were published in the U.S. matched by publication year, primary USPC technology class, and the number of citations up to 1949. To classify a patent by its most likely industry, we use the data of Kerr (2008). We assign to each USPC class the most likely four-digit SIC industry in which it is used. A solid circle indicates that a coefficient is significant at the 10% level. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Foreclosure seems to be particular harmful for start-up innovation. In columns (7) and (8) of Table 1.3 we show that small and young companies increased their citations only outside the field of telecommunications, but not inside.[36] As a large part of the effect in the full sample was driven by small and young companies, this suggests that also start-ups react strongly to market foreclosure. In fields outside of telecommunications, compulsory licensing fostered innovation by small and young companies since Bell as the supplier of technology did not control product markets through vertical integration or via exclusive contracts.

Our results suggest that market foreclosure stifles innovation and that compulsory licensing is not sufficient to foster innovation without supporting structural remedies. This confirms the general perception at the time of the lawsuit. Both the public and antitrust officials were aware that because of Bell's persistent monopoly, compulsory licensing would only help companies outside the telecommunications field. A witness in the Congressional hearings put it succinctly: "while patents are made available to independent equipment manufacturers, no market for telephone equipment is supplied (...). It is rather a useless thing to be permitted to manufacture under patent if there is no market in which you can sell the product on which the patent is based." The Antitrust Subcommittee concluded that "[t]he patent and technical information requirement have efficacy only so far as they permit independent manufacturers to avail themselves of patents in fields that are unrelated to the common carrier communication business carried on by the Bell System companies, and nothing more." On May 4, 1954, presiding Judge Stanley N. Barnes suggested that compulsory licensing policy for reasonable rates is "only good window dressing" but would do no good because Western Electric had already "achieved an exclusive position (...) and liberal licensing would not permit competitors to catch up" in the telecommunications business (Antitrust Subcommittee, 1959, pp. 108).

In the years after the consent decree, the Bell System faced repeated allegations of exclusionary behavior. By the 1960s and 1970s, a range of new firms were eager to enter the telecommunications market but Bell implemented measures to make it expensive or impossible (Wu, 2012). This led to a number of regulatory actions, for example forcing interconnections of Bell's telephone system to the entering competitors MCI in 1971 which

---

[36]We use the most likely SIC code to determine the field of the citing patent.

provided long distance services using microwave towers (Temin and Galambos, 1987; Gertner, 2012, p. 272). Eventually, the continued monopolization of the telecommunications market by Bell resulted in the 1974 antitrust lawsuits. The lawsuit mirrored almost scene by scene the case of 1949. Again, Bell was charged with excluding competitors from the market of telecommunications equipment. And again, the Department of Defense intervened on the grounds of national defense. But the Reagan administration was not as accommodating as the Eisenhower administration had been and the Department of Justice was keen on going after Bell. The case ended with the break-up of the Bell System in 1983, opening up the market for telecommunications equipment for competition.

1.6 The Consent Decree Increased U.S. Innovation in the Long Run

The historical set-up of the Bell case gives us the opportunity to look also at the long-run innovation effects of a consent decree. In the previous section, we showed that the increase in follow-on citations is measurable for the first five years. This raises the question how lasting the impact of a large-scale intervention in patent rights really is. To answer this question, we study the long-run impact of the case against Bell on the patent activities of firms patenting in the U.S. More specifically, we examine the increase in the total number of patents in a USPC technology subclass with a compulsorily licensed Bell patent relative to a subclass without. We employ the following empirical model

$$\#Patents_{s,t} = \beta_t \cdot I(Bell > 0)_s + Controls + \varepsilon_{s,t} \tag{1.3}$$

where the outcome variable is the total number of patents in a technology subclass s (Moser and Voena, 2012; Moser et al., 2014). The treatment variable equals one if there is at least one compulsorily licensed patent in the technology subclass. As controls, we use USPC class-year fixed effects.[37] Our sample consists of 235 classes with 6'276 subclasses

---

[37]To follow the literature we use USPC technology classes here and not SIC classes.

of which 1'209 are treated.[38]

In Figure 1.10a we plot the number of excess patents for all patent classes. We leave out patents by Bell to focus on patenting of other companies. Starting in 1953, the number of patents in technology classes where Bell patents were compulsorily licensed increased relative to subclasses without Bell patents, and it continued to do so beyond 1960, when the last Bell patents affected by the consent decree expired. This suggests that the consent decree increased U.S. innovation in the long run.

To quantify the effect, we next estimate the average yearly effect of the consent decree on the total number of patent applications for the time period 1949-1960. We employ the following difference-in-differences model:

$$\#Patents_{s,t} = \beta_1 \cdot I(Bell > 0)_s + \beta_2 \cdot I(Bell > 0)_s \cdot I[1955 - 1960] + Controls + \varepsilon_{s,t} \quad (1.4)$$

where $I(Bell > 0)_s$ is 1 if Bell has a patent in the subcategory s. As controls we use class-year fixed effects.

The coefficients are reported in Table 1.4. In the first five years alone, patent applications increased by 2.5 patent applications in treated classes (column 1). This is an increase of around 24.5%.[39] Furthermore, patent applications by new companies entering the market increased relatively more than patent applications by other companies (columns 2 and 3).[40]

The increase appears to be stronger outside of telecommunications technologies (column 4 and 5). In Figure 1.10b we plot the average treatment effects estimated with Equation 1.4 for different levels of closeness to the production of telecommunications equipment. Again the effects are weak for technologies closely related to the production of telecom-

---

[38]We exclude subclasses that did not have any patents at all before 1956 and we include only patent classes that contain subclasses that were treated and subclasses that were not.

[39]Untreated subclasses have on average 2.17 patent applications in the the pre-treatment period. In these subclasses the number of patent applications increase by 0.52 from the pre- to the post-treatment period. Using the estimate for the difference between treated and untreated classes, 7.5, in column (1) of 1.4, we calculate the counterfactual number of applications in treated classes in the absence of compulsory licensing which is equal to 10.19 (=7.5 + 2.17+0.52). The treatment effect is 2.5. Thus, the number of patents increased relative to the counterfactual by 24.5% (=2.5/10.19).

[40]The number of patents of young and small increases by 38% while the number of patents of all other companies increases by 18%.

Figure 1.10: Impact of the Consent Decree in the Long Run

(a) Annual Treatment Effects on the Number of Patent Applications
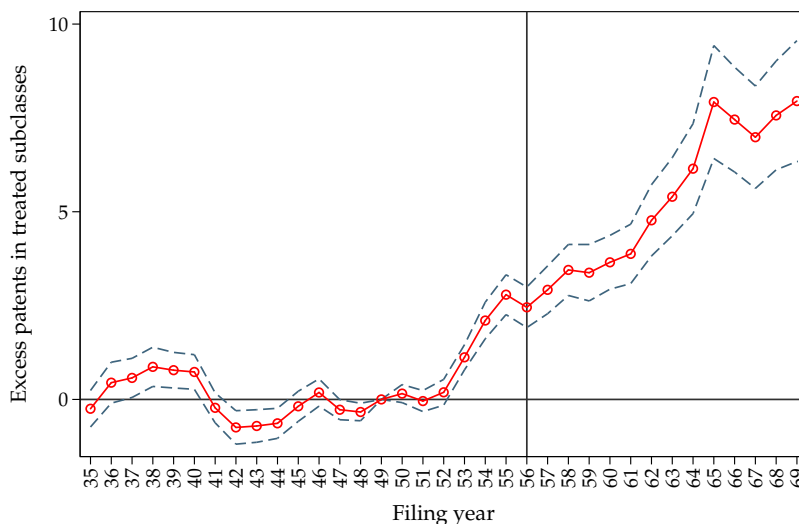
(b) Excess Patents sorted by Likelihood of Being Used in the Production of Telecommunications Equipment
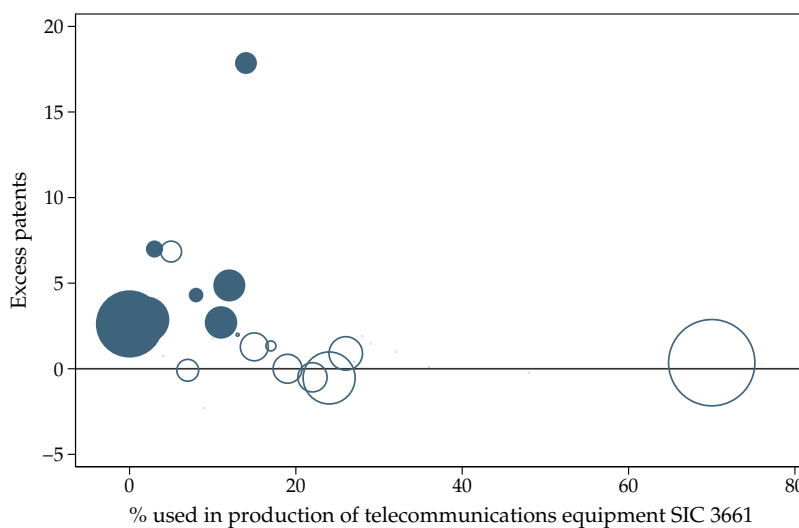
Notes: The dependent variable is the number of patent applications per (aggregated) subclass per year. A subclass is treated if it contains at least one Bell patent that was subject to compulsory licensing. Sub-figure (a) shows annual treatment effects β_t estimated with Equation 1.3 for all patent classes. Standard errors are clustered at the class level. Subfigure (b) shows the average increase in the number of patents β_2 estimated with Equation 1.4 for patent classes with varying likelihood of being used in the production of telecommunications equipment. To determine the likelihood that a patent is used in industry SIC 3661 we use the data of Kerr (2008). The size of the circle signifies the number of Bell patents in a technology and a solid circle implies that the coefficient is significant on the 10% level. The data are from from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

munications equipment and strong for unrelated technologies. This again suggests that the fields in which Bell continued to operate experienced slower technological progress than markets where entry of start-ups was possible.[41]

Figure 1.10a shows that the increase in patenting begins in 1953, two years before the increase in citations to Bell patents. In 1953, Bell's most important invention, the transistor, became available for licensing, spurring the creation of the computer industry. To make sure that the entire increase is not driven by this one exceptional invention, we analyze computer and non-computer patents separately and report the results in columns (6) and (7) of Table 1.4. The effect is stronger for the computer patents, but the increase in patenting is also significant without any computer patent.

Thus, overall we find that the consent decree led to a long-lasting increase in the scale of innovation mainly outside the telecommunications field. This is consistent with the theoretical argument by Acemoglu and Akcigit who build on the the step-by-step innovation model of Aghion et al. (2001) to analyze the effects of compulsory licensing on innovation (Acemoglu and Akcigit, 2012). They consider the case where all current and future patents in the economy are compulsorily licensed for a positive price and identify two main effects. On the one hand, compulsory licensing helps technological laggards to catch up and brings more industries to a state of intense competition. This "composition effect" increases innovation, because companies in industries with intense competition invest more in R&D in order to become the industry leader. On the other hand, compulsory licensing reduces the time a technology leader keeps its profitable position. This "disincentive effect" reduces innovation and growth in the economy.

In our case, compulsory licensing was selectively applied to one company that did not participate in any market other than the telecommunications market. This enabled many new companies to enter markets with state-of-the art technology and to compete for the industry leadership with full patent protection of future inventions intact (Holbrook et al., 2000).

---

[41]In unreported regressions we use citation-weighted patents instead of the absolute number of patents and find the same results. Results are available from the authors upon request.

Table 1.4: Patent Applications per Subclass and Year by Company Type and Field

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Type of Company			Field		
	Baseline	Young & Small	Others	Communication	Not Comm.	Computer	Not Computer
Treated	7.5***	2.2***	5.3***	5.2***	8.2***	4.4***	7.7***
	(0.6)	(0.2)	(0.4)	(0.3)	(0.7)	(0.4)	(0.6)
Treated x I(55-60)	2.5***	1.2***	1.3***	1.4***	2.7***	3.0***	2.5***
	(0.3)	(0.1)	(0.2)	(0.5)	(0.3)	(0.8)	(0.3)
Clusters	235	235	235		211		222
Observations	75312	75312	75312	6228	69084	2268	73044

Notes: The dependent variable is the number of patent applications per subclass per year. A subclass is in the treatment group if it contains at least one Bell patent that was subject to compulsory licensing. This treatment variable is interacted with an indicator that is 1 for the period 1955-1960. The panel 1949-1960 includes subclasses that had at least one patent application between 1940 and 1949 and whose classes contain both treated and untreated subclasses. Young companies are companies with their first patent granted less than ten years ago and small companies are companies with less than 10 patents in 1949. We classify a technology as related to communication if its most likely industry is SIC 3661, the production of telecommunications equipment. The last column excludes all computer patents, defined as patents that belong to the NBER technological subcategory Computer Hardware and Software, Computer Peripherals and Information Storage (Hall et al., 2001). The patent data are from the Worldwide Patent Statistical Database (PAT-STAT) of the European Patent Office. The regressions include class-year fixed effects and standard errors are clustered at the class level, except in columns (4) and (6), where the sample is limited to 24 classes in telecommunications and 13 in computers, respectively. *, **, *** denote statistical significance on 10%, 5% and 1% level, respectively.

Thus, in all industries but the telecommunications industry we measure the pure composition effect without the counteracting disincentive effect. The interpretation that the consent decree helped to open up new markets and enabled new start-ups to compete is consistent with historical accounts on the growth of electronics and computers industry in the 1950s and 1960s (Grindley and Teece, 1997).

1.7 Case Study: The Diffusion of the Transistor Technology

In this section, we examine the diffusion of the transistor technology because it is a particularly insightful case study for the mechanisms illustrated in the previous sections for three reasons: First, in response to the antitrust lawsuit Bell already started in 1952 to license the transistor technology via standardized non-discriminatory licensing contracts. This creates an interesting variation in the timing of licensing. Second, transistor patents were expected to be particularly important, hence we can estimate how the amount of follow-on innovation varies with patent quality. And finally, under the impression of the antitrust lawsuit Bell was very careful not to engage in exclusionary practices with its transistor patents. Thus, in 1956 the only change for the transistor technology was that the patents were now royalty free. This allows us to examine the isolated impact of a decrease in royalties.

The transistor is arguably the most important invention of Bell Labs. As the most basic element of modern computers, the transistor has been instrumental in the creation of entire industries and its invention heralded the beginning of the information age. The invention of the transistor earned John Bardeen, Walter Brattain, and William Shockley the Nobel Prize in Physics in 1956. They filed patents in June 1948 and announced the invention on July 1 of the same year. The patents were published in 1950 and 1951. Bell, the military, and the research community at large immediately understood the importance and value of the transistor.

Due to the ongoing antitrust lawsuit, Bell's management was reluctant to draw attention to its market power by charging high prices for transistor components or for licenses

(Mowery, 2011). To appease the regulator, Bell's top managers agreed to share and license the transistor device with standardized non-discriminatory licensing contracts (Gertner, 2012, p.111). In addition, Bell decided to actively promote the transistor by organizing conferences to explain the technology. In April 1952, over 100 representatives from 40 companies gathered for a nine-day Transistor Technology Symposium, including a visit to Western Electric's transistor manufacturing plant in Allentown, PA. After the conference, 30 companies decided to license the transistor technology for a non-refundable advance payment of \$25'000 ($\sim$ \$220'000 in today's dollars) that was credited against future royalty payments (Antitrust Subcommittee, 1958, p.2957). Royalty rates amounted to 5% of the net selling price of the transistor in 1950, which were reduced to 2% in 1953 (Antitrust Subcommittee, 1959, p. 117).

To be able to separately analyze the transistor we identify among the patents affected by the consent decree all patents related to the original transistor inventor team. There are two main transistor patents: Patent # 2,524,035 with the title "Three-Electrode Circuit Element Utilizing Semiconductive Materials" granted in 1950 to John Bardeen and Walter Brattain and Patent # 2,569,347 with the title "Circuit Element Utilizing Semiconductive Material" issued to William Shockley in 1951. To these two patents, we add all the patents of all researchers who actively worked towards the development of the transistor at Bell Labs.[42] Then we add all patents from all co-authors. We identify 329 "transistor" patents affected by the consent decree (i.e., held by Bell Labs). This sample is most likely a super-set of all transistor patents. For example, it also includes patent # 2,402,662 with the title "Light Sensitive Device" granted to Russell Ohl, the original patent of the solar cell. The median publication year of the patents in the transistor subsample is 1947; and 168 of these patents are also included in our baseline sample.

To be able to repeat our regressions in this subsample of transistor patents we extend our baseline sample to patents published up to 1951. As control group, we now use patents with the same number of pre-citations up to 1951 while all other criteria stay the same.

Figure 1.11 shows the yearly excess citations of transistor patents relative to control group

---

[42]Researchers whom we classify to have actively contributed to the transistor at Bell Labs were in alphabetical order Bardeen, Bown, Brattain, Fletcher, Gardner Pfann, Gibney, Pearson, Morgan, Ohl, Scaff, Shockley, Sparks, Teal and Theurer (e.g. Nelson, 1962).

Figure 1.11: Annual Treatment Effects on Excess Citations of Transistor Patents
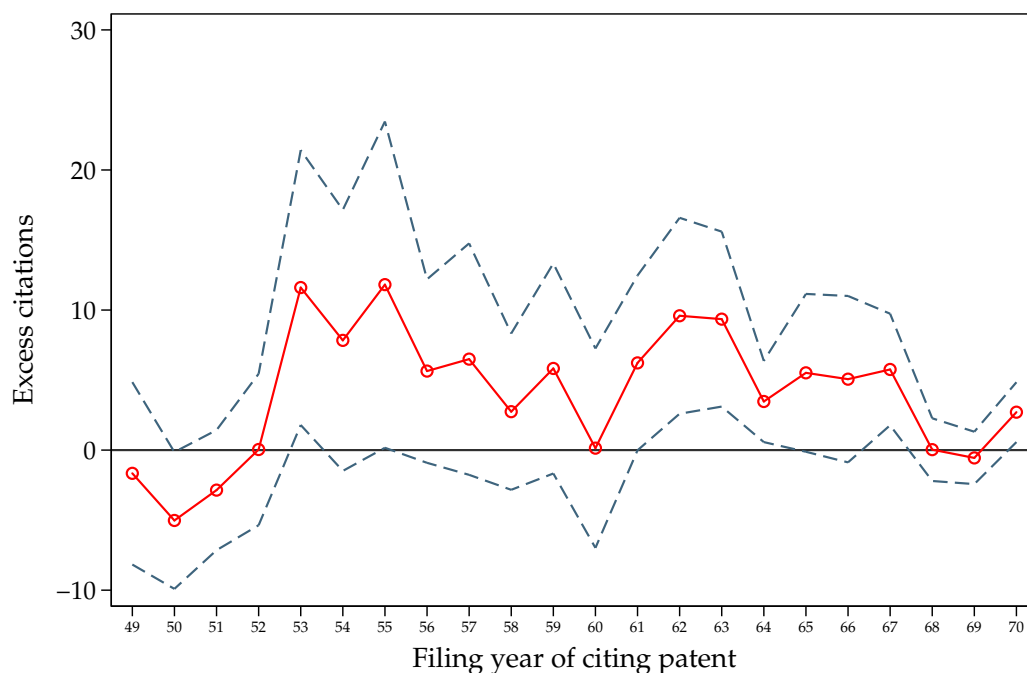
Notes: This graph shows the estimated number of yearly excess citations of transistor-related patents affected by the consent decree ("Bell patents") relative to patents with the same publication year, in the same three-digit USPC primary class and with the same number of citations up to 1951. We define Bell patents as transistor-related if they are either one of the two main transistor patents (Patent # 2,524,035 or Patent # 2,569,347) or were filed by inventors associated with these patents or their co-inventors. To arrive at these estimates, we regress the number of citations in each year on an indicator variable that is equal to one if the patent under consideration is affected by the consent decree and year fixed effects. The dashed lines represent the 90% confidence bands for the estimated coefficient. To adjust for the different number of control patents per treatment patent in each stratum, we use the weights suggested by Iacus et al. (2009). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

patents. The coefficient of 1952, which is not matched and is close to zero, speaks in favor of parallel trends. The impact of licensing is measurable starting in 1953, and lasts for at least 15 years. This suggests that standardized licensing had a positive impact on follow-on innovation. The fact that the impact does not strongly increase in 1956 when the consent decree reduced licensing fees to zero suggests instead that the price effect of compulsory licensing had little further impact.

Table 1.5: The Transistor Subsample

	(1)	(2)	(3)	(4)	(5)	(6)
	Publication year <1952			Publication year<1949		
Subsample	Baseline	Transistor	No transistor	Baseline	Transistor	No transistor
Start treatment	1955	1953	1955	1955	1953	1955
Bell	-0.3	-1.4	-0.4	-0.4	-0.9	-0.4
	(0.3)	(1.2)	(0.3)	(0.5)	(2.1)	(0.5)
I(53/55-60)	-5.7***	-6.3**	-5.6***	-6.4***	-7.4***	-6.4***
	(0.7)	(2.7)	(0.7)	(0.6)	(2.2)	(0.6)
Bell x I(53/55-60)	1.9***	8.0**	1.8***	2.0***	4.4*	2.0***
	(0.5)	(3.7)	(0.5)	(0.6)	(2.3)	(0.6)
Constant	19.0***	23.0***	18.8***	18.3***	22.3***	18.1***
	(1.4)	(3.2)	(1.4)	(1.2)	(2.9)	(1.2)
# treated	5758	204	5554	4533	168	4365
Clusters	239	65	237	225	58	223
Obs.	1035421	64891	1021733	896556	56664	886044

Notes: This table shows the results from a difference-in-differences estimation. As the dependent variable we use all citations by companies other than the filing company. For the regression with the transistor patents, we define the treatment period as starting in 1953; for the non-transistor patents we define the treatment period as starting in 1955, as in our main regression in equation (2). *Bell* is an indicator variable equal to one if a patent is published by a Bell System company before 1949 and is therefore affected by the consent decree. As control patents, we use all patents that were published in the U.S. matched by publication year, primary USPC technology class, and the number of citations. We define patents as transistor patents if they were filed by a member of the original transistor team or one of their co-authors. In the regressions for columns (1) to (3), we use all patents with a publication year before 1952 and we match all citations up to and including 1951. Correspondingly, in the regressions for columns (4) to (6) we use patents and citations up to 1949. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. All coefficients are multiplied by 100 for better readability. Standard errors are clustered on the primary three-digit USPC technology class level and *, **, *** denote statistical significance on 10%, 5% and 1% level, respectively.

Table 1.5 reports the results from repeating our baseline regression in this subsample. We find that citations to the transistor patents increase by 52% (column 2). They experience around a four times higher increase in follow-on citations than other consent decree patents. The magnitude of the effect is consistent with the presumption that patents on more important inventions experience a larger increase after compulsory licensing.

Despite the large effects, the transistor patents do not drive the effect in our main sample. To rule out this possibility we analyze our original sample up to 1949 with and without transistor patents. Results are shown in columns (5) and (6). We find large but barely significant effects for the transistor sample and virtually the same effect without transistors as in the baseline regression that includes transistor patents (column 4).[43]

Transistors are the classical example of a general purpose technology that has the potential of having a large scale impact on the economy (Helpman, 1998). If it had not been for the antitrust lawsuit against Bell, odds are that Bell's licensing policy would have been less accommodating and the follow on-innovations stimulated by the transistors less dramatic than they were.

1.8   Conclusion

In this chapter, we show that antitrust enforcement can increase innovation. The 1956 consent decree that settled the antitrust lawsuit against Bell increased innovation, mostly by small and young companies building on Bell's established technologies. We conclude that antitrust enforcement can play an important role in increasing innovation by facilitating market entry.

Several antitrust scholars have argued that antitrust enforcement should pay special attention to exclusionary practices because of their negative influence on innovation (Baker, 2012; Wu, 2012). Our study seconds this view. We show that foreclosure has a negative impact on innovation and that compulsory licensing may not be an effective remedy to end market foreclosure and to overcome its stifling effect on innovation unless accompanied by structural remedies.

---

[43]The large magnitude of the effect should not be taken at face value. The identifying assumption of this regression is that the control patents would have had the same number of citations as the transistor patents. In our regression this is true for 1953, but given the exceptional nature of the invention of the transistor, it is fair to assume that this trend might have diverged in later years. Furthermore, it is not absolutely clear from the historical records why Bell decided to license the transistor patents. If the licensing decision was taken because of the expectation of important follow-on research, our estimate might give an upper bound on the effect. For example, Jack Morton, the leader of Bell Labs effort to produce transistors at scale, advocated the sharing of the transistor to benefit from advances made elsewhere. Source: http://www.computerhistory.org/siliconengine/bell-labs-licenses-transistor-technology/ (last accessed September 09, 2016).

Compulsory licensing is often imposed in merger cases where the market structure changes endogenously (Delrahim, 2004; Sturiale, 2011). We would expect that if the newly merged company is able to foreclose the product market, compulsory licensing is not an effective remedy. More empirical studies are needed to assess whether the negative effect of market foreclosure on innovation is a first order concern for merger and acquisition cases.

We estimate the negative effects of patents on follow-on innovations by other companies, but we cannot determine how large the incentive effect of patents for the company holding the patent is. In our case, compulsory licensing does not appear to have had a strong negative effect on Bell's patenting activities. It would be surprising if this was the norm (Williams, 2015). But it is consistent with Galasso and Schankerman (forthcoming) who show that large companies do not reduce their innovation activity when their patents are invalidated in court, but do change the direction of their research and development activities.

We analyze a very important antitrust lawsuit from the 1950s. Using a historical setting has the advantage that we can draw on a large number of detailed historical accounts and that we can conduct a long run evaluation many years after the case. At the same time it is unclear whether the size of the effects of compulsory licensing would be similar today. Jaffe and Lerner (2011) suggest that many negative effects of the patent system discussed today are related to regulatory changes surrounding the establishment of the Court of Appeals for the Federal Circuit in 1982. The reforms led to a significant broadening and strengthening of the rights of patent holders and consequently to a surge in the number of patents granted. This makes us think that the effects of compulsory licensing might be even larger today.

Chapter 2

Disclosure and Cumulative Innovation: Evidence from the Patent Depository Library Program

2.1 Introduction

How does information disclosure through patents affect innovation? Disclosure of technical information is often invoked as one of the patent systems' key economic justifications (e.g., Scotchmer and Green, 1990). Courts and intellectual property lawyers also emphasize the importance of disclosure for innovation while academic observers doubt the effectiveness of the requirement (Roin, 2005). In practice, in patent laws around the world inadequate disclosure of underlying technical information can invalidate patent rights (Hall and Harhoff, 2012). So far, there is little empirical evidence on the actual impact of disclosure on follow-on innovation.

The reason is that while the question is central in the study of cumulative innovation, it is particularly challenging to analyze empirically (Graham and Hegde, 2015; Hegde and Luo, 2017; Williams, 2017). Since the patent system gives the right to exclude competitors only conditional on the disclosure of technical information, there is little variation that would enable researchers to identify the "enablement effect" of disclosure, i.e., the value of

---

This chapter is based on joint work with Jeffrey Furman and Martin Watzinger (Furman et al., 2017).

information provision on subsequent innovation, separately from the effects of exclusion.

In this chapter, we take advantage of the expansion of the USPTO Patent and Trademark Depository Library (PTDL or patent libraries) system from 1977 to 1997 to investigate the effect of disclosure of patent information on innovation in the regions around the newly-created patent libraries. While the exclusion rights associated with patents are national in scope, the opening of these patent libraries in a period before the Internet yielded regional variation in the cost to access the technical information disclosed in patent documents.

This patent library system was created in the 1800s to provide patents and innovation-related resources for independent inventors, entrepreneurs, and incumbent firms. By 1977, 22 libraries had been established, primarily in New England and West of the Mississippi. Beginning in 1977 and continuing until 1997, the USPTO embarked on an effort to open at least one patent library in each of the U.S. states. These libraries provided information on granted patents, including search tools, which were otherwise not available to inventors or attorneys outside of the PTO headquarters location in Washington, DC, or in other patent libraries.

Whereas the location of libraries prior to 1977 had been chosen by the PTO, libraries in the 1977 to 1997 period were granted with some degree of randomness within each state and in its time of opening, as they were typically given to the first qualified library in a state to request them. While some major cities in a state (e.g., Boston, MA) were more likely to have the capacity and demand for such institutions than smaller cities or towns (e.g., Springfield, MA or Worcester, MA), it is not clear whether innovation trends are more likely to have driven requests for such libraries in cities of similar sizes within states (e.g., Kansas City, MO vs. St. Louis, MO). We leverage this dimension of randomness in location and timing to estimate the impact of disclosure on subsequent innovation.

In our main specification, we compare the number of patents in the close vicinity around the patent library with the change in the number of patents around Federal Depository Libraries (FDLs). The 1'252 Federal Depository Libraries make government documents such as laws and Acts of Congress freely available to the public. As the missions of patent libraries and FDLs are similar - provide the public with official documents - almost all patent libraries are also Federal Depository Libraries. Patent libraries are usually first an

FDL and only later become patent libraries. The reason is that according to one librarian, "a factor that would influence a library in becoming a patent library is whether they had been involved with government documents in another capacity". For each patent library, we use all Federal Depository Libraries that are located in the same state and within 250 km as a control group.

We find that after a patent library opens the number of patents within 25 km increases by on average around two patents per year or 18% relative to the pre-opening mean. This effect is localized and becomes insignificant at more than 25 km. In line with increased access to patents driving this effect, we find that young and small companies which plausibly face larger barriers to access than large companies increase patenting more. The increase in patenting is most pronounced if the patent library is also a university library. This result suggests that there is a complementarity between access to patent knowledge and technical education for the production of innovation.

We show that it is unlikely that concurrent shocks drive these effects. In the years before the opening, the number of patents around the control libraries are identical compared to the soon to be designated patent libraries. This speaks in favor of parallel trends. There is also no differential trend between control libraries suggesting that the libraries not only relocate innovative activities from close-by regions. Our results are robust to the use of a closer or a looser control group.

In additional analyses, we also find that the structure of patents changes after a patent library opens: In particular, the distance to patents cited by inventors living close to a patent library increases. Apparently, after a patent library opens inventors start to work on problems that are less local and the geography of innovation becomes more dispersed. We do not find substantial evidence that patents start to cite different technological fields or that the patents filed after the patent library opened are of higher quality. Thus, access to prior art facilitates the construction of an 'invisible college' of like-minded inventors building on each others' ideas rather than induce inventors to work on different sorts of problems.

This study is the first to show that access to technical information disclosed in patents can increase innovation. Disclosure is thought to be one of the key functions of the

patent system. For example, Machlup (1958) argues that a patent "serves to disseminate technological information, and that this accelerates the growth of productivity in the economy". Yet, critics argue that the usefulness of disclosure is limited and several inventor surveys find no or only modest benefits from reading patents (Cohen et al., 2002; Arora et al., 2008; Gambardella et al., 2011; Hall and Harhoff, 2012).[1] Newer studies on the American Inventor Protection Act find that many inventors voluntarily disclose their invention, leading to earlier licensing deals (Graham and Hegde, 2015; Hegde and Luo, 2017). Our study adds to this literature by showing that increased access to technical information provided by patent libraries increases patenting for the subsample of small and young companies.

More generally, our study enhances our understanding of the role of research enhancing institutions by showing that investment in patent libraries helped to fuel regional innovation. Research enhancing institutions lower the costs of access to useful knowledge and thus help to foster geographical and intertemporal spillovers on which economic growth is based (Mokyr, 2002). In related work, Biasi and Moser (2016) find that reducing the access costs to science books during World War I increased scientific output in particular in regions with libraries buying these books. Close to our work, Furman and Stern (2011) show that a biological resource center, a library of living organisms, helps to foster follow-on innovation because it provides input for the research process. Our findings contribute to earlier research by showing that patent libraries increased innovation across U.S. states by improving access to technical information.

The remainder of this chapter is organized as follows. Section 2.2 describes the U.S. Patent Depository Library Program and the U.S. Federal Depository Library Program. In Section 2.3 we describe the data and the empirical strategy. In Section 2.4 we show that opening a patent library increased innovation in its close vicinity, we examine the heterogeneity of the effects for different firm types, and we present auxiliary results and robustness checks. Section 2.5 looks at the changing citation behavior of inventors and shows that inventors start to cite more distant work. Section 2.6 concludes.

---

[1] Inventions that are hard to reverse engineer are more effectively protected by secrecy and thus not disclosed. In addition, patents are often willfully opaque and there is a legal risk in the U.S. from using patents as source of information because it increases the likelihood of being found to infringe on the patent.

2.2   The U.S. Patent Depository Library Program

The nationwide network of Patent and Trademark Depository Libraries (PTDLs) traces its beginning to the year 1871 when the United States Patent and Trademark Office (USPTO), then known as The Patent Office, first started distributing copies of patents to a small number of libraries.[2] Until then, patent documents were housed at only one location, at The Patent Office, in Washington, DC.[3]

It took an act of congress, 35 U.S.C. 13, to enable the Patent Office to start disseminating patent information to the public. Libraries were given the option of serving as a patent depository, and eight libraries received patent documents in 1871: The New York State Library, the Boston Public Library, The Public Library of Cincinnati and Hamilton County, the Science and Engineering Library at Ohio State University, the Detroit Public Library, the Los Angeles Public Library, the New York Public Library, and The St. Louis Public Library. By 1977, the number of patent libraries had grown to 22 of which 14 became patent libraries in the 19th century. Most of the original libraries were very large public libraries located in the industrial midwest and eastern seaboard and most were east of the Mississippi River. They received unbound paper patents on a weekly basis along with the Official Gazette of the U.S. Patent and Trademark Office and two search indices.

After 100 years of relative inactivity, the year 1977 marked the start of a substantial expansion of the patent library system with the aim to increase the number of patent libraries by at least three each year and to put at least one patent library in each state.[4] The goal to put a patent library in every state was achieved twenty years later in 1997. The map in Figure 2.1a shows the 22 libraries that were designated before 1977 and figure 2.1b shows all 84 patent libraries which opened until 2017 in the continental U.S. Table B.1 and Table B.2 in appendix B list patent libraries up to 2002. Currently, about half

---

[2]This section follows the description in Sneed (1998) and Jenda (2005).

[3]When the publication of the Official Gazette, a weekly publication of the USPTO that lists patent abstracts and a representative drawing of the invention, started in 1872, this title was added to the list of documents that were distributed to libraries.

[4]The reason was USPTO assistant commissioner William I. Merkin who started to evaluate the patent library system in 1974 and led its overhaul with a series of conferences beginning in April 1977.

of the membership consists of academic libraries with nearly as many public libraries.[5] After 1997, the patent library system adopted a new goal of controlled growth in areas of high population combined with high patent and trademark activity which warrant the resources invested by the USPTO (Sneed, 2000).

Starting in 1977, the patent library system was not only expanded but also reorganized: Libraries could apply to become patent libraries if they fulfilled a number of requirements. First, they had to acquire a collection of all U.S. utility patents issued 20 years prior to the date of designation. Second, each patent library had to have trained staff to assist the public in the search for prior art. To ensure adequate training each patent library must have a representative at every annual PTDL Training Seminar in Washington, DC. Third, they had to provide free public access and a collection of search tools for the public.

Several librarians that we contacted stated that it is necessary to be a large library in order to be able to fulfill the resource intensive requirements to become a Patent Depository Library. This includes the space to host the patent documents, the availability of qualified staff and resources to meet the start-up, and ongoing costs. In later years, the space requirement became less a concern after the introduction of microfilm. Indeed the conversion from paper to microfilm distribution has been cited as a reason why many libraries joined the program after 1982.

Usually a patent library already had a history of handling government documents as Federal Depository Library, according to the interviewed librarians. In our sample, 86% of patent libraries are part of the Federal Depository Library Program. Federal Depository Libraries make U.S. federal government publications available to the public at no cost. As of 2008, there were 1'252 Federal Depository Libraries, at least two in each of the 435 Congressional Districts. There are two ways in which a library may qualify for FDL status: First, each member of Congress may delegate two qualified libraries or a library may be designated. Second, libraries at land-grant colleges and universities, libraries of federal agencies, the highest appellate court of a state, or accredited law schools automatically qualify for the status of Federal Depository Library. Because of this structure and the attached requirements to serve as a library in either program, Federal Depository Libraries

---

[5]Since 1871, six PTDLs have withdrawn for various reasons, including library closing, no funding for the back file, and a change in institutional priority creating a lack of ability to perform the service.

Figure 2.1: Location of all Patent Libraries in the U.S.

(a) Up to 1977
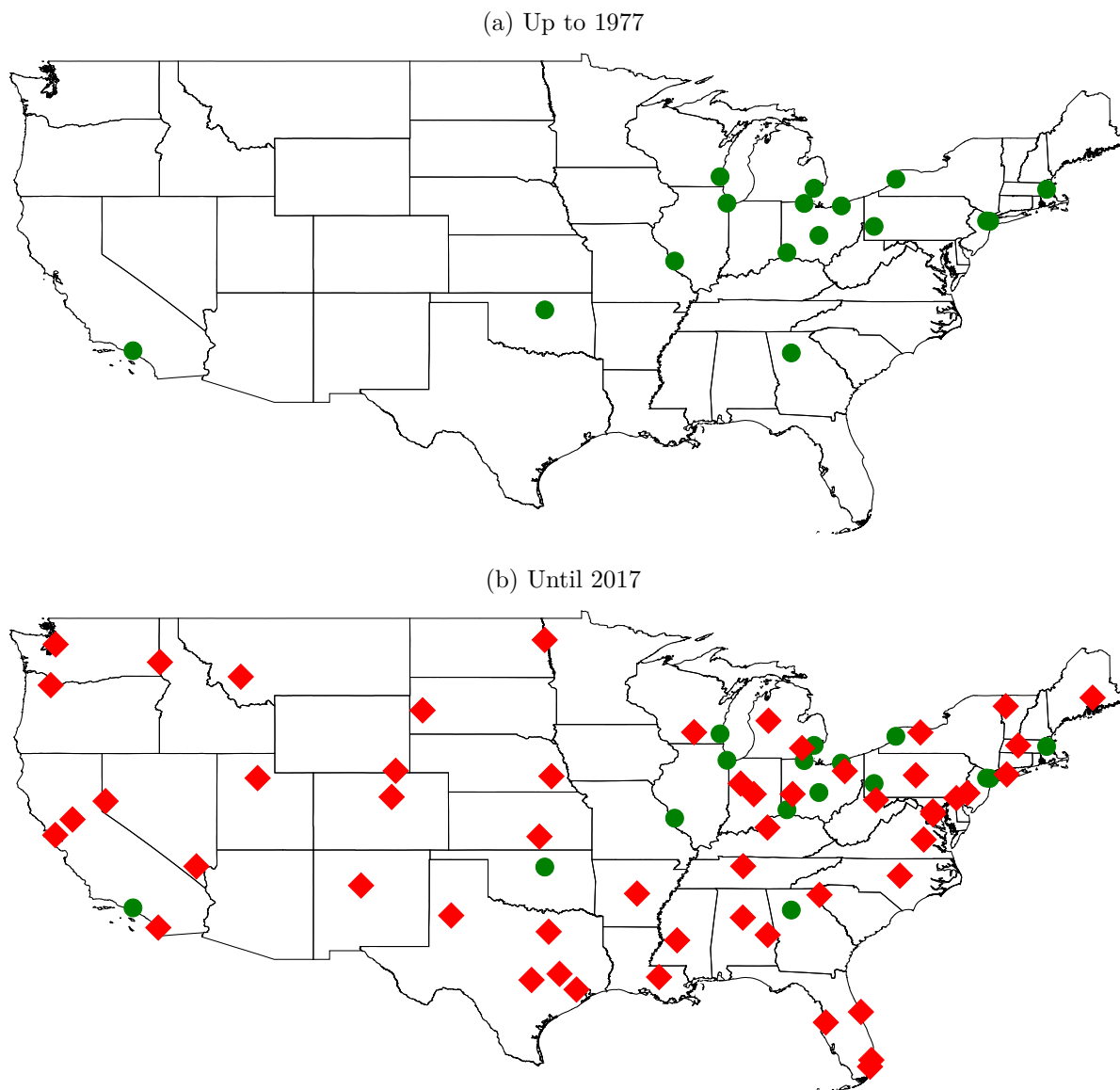
(b) Until 2017

Notes: The green circles indicate patent libraries in the continental United States opened before the major expansion in 1977. The red diamonds show the location of patent libraries opened in or after 1977 until 2017.

Figure 2.2: Patent Libraries and Control Libraries
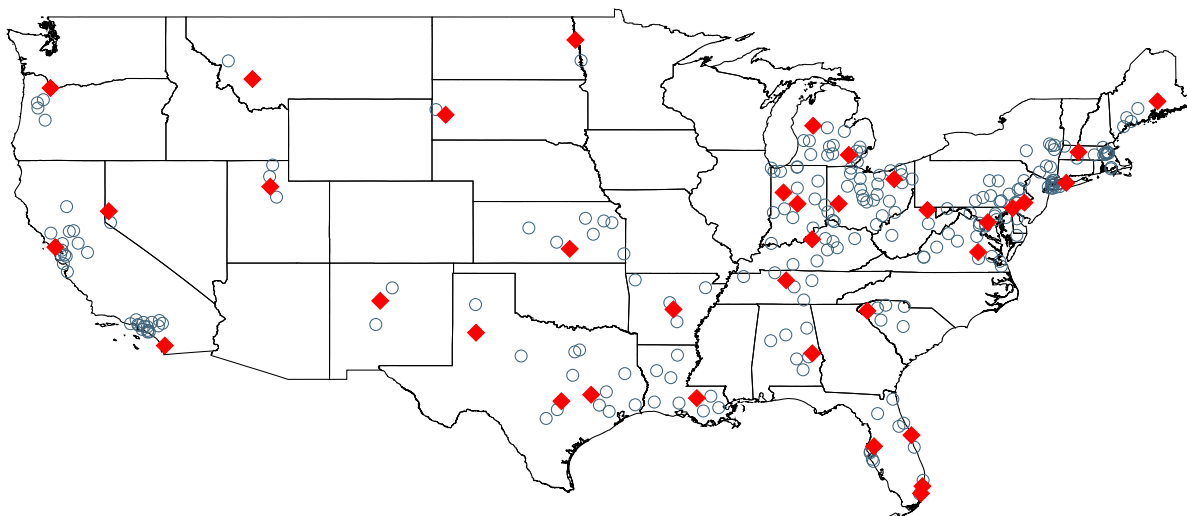
Notes: The red dots show the locations of patent libraries in our main sample. The blue hollow dots show the locations of corresponding control libraries.

are a natural control group for Patent Depository Libraries.

2.3 Data and Empirical Setup

For our empirical analysis, we assemble a dataset of all patent and federal depository libraries in the U.S. The data on the opening dates of each patent library is taken from Jenda (2005) and the complete list of Federal Depository Libraries is from the online Federal Depository Library Directory.[6] We geolocate all patents by city and state and merge it with population data from the U.S. census.

We combine the data on libraries with patent data from the Harvard project on patent inventor disambiguation as described in Li et al. (2014). This data covers the time period between 1975 and 2010 and is based on data from the USPTO and the NBER. The dataset includes patent information and information on inventors. For each patent, we compute backward and forward citations along with measures for the generality and originality of cited and citing patents (Hall et al., 2001).

As outcome, we use the number of patents within 25 km of the new library to measure

---

[6]The Federal Depository Library Directory is available on https://catalog.gpo.gov/fdlpdir/FDLPdir.jsp (last accessed 2017-07-30).

the effect of the opening of a patent library on innovation in the area. We normalize the number of patents by the population in the area to adjust for different city sizes. Using patents as a measure for innovative output is standard but not uncontroversial. In our particular case, patent libraries could also increase patenting without increasing innovation because they might make it easier to file a patent or because the librarians may give advice on how to structure a patent. Yet, this seems unlikely because a U.S. patent application can be mailed from any post office and the employees of patent libraries are only allowed to help with the search for prior art but not with the preparation of a patent filing. Additionally, because we want to measure real innovation, we delete those patents which are never cited.

To construct a counterfactual for the patent libraries, we use as control group Federal Depository Libraries in the same state and within 250 km (but not within 25km) in our main specification. Federal Depository Libraries provide access to governmental information to each citizen. As of 2017, there are 1'251 Federal Depository Libraries in the U.S., with at least two in each Congressional District. Most patent libraries are Federal Depository Libraries that chose to become Patent Depository Libraries. 84% of all patent libraries are Federal Depository Libraries and 96% of them are classified as "medium" or "large libraries". To make use of this setup, we assign to each patent library that is also an FDL, all other medium or large FDLs within the same state and 250 km as control group.

We finally drop all patent libraries without a control library, the patent libraries in Hawaii, Puerto Rico and all other U.S. overseas territory, and all patent libraries that opened before our data started.[7] In our main sample, we also drop the libaries of Rochester in the state of New York and Burlington in Vermont. The reason is that both have an extremely high patent per capita ratio because they host Kodak, Xerox, and Bausch & Lomb in case of Rochester and IBM in case of Burlington. As a consequence we cannot find a valid control group for these two libraries. Thus we arrive at 38 patent libraries that opened after 1979 along with 415 control libraries.[8] Figure 2.2 shows the position of

---

[7]Our required five year pre-opening period in fact implies that we drop libraries which opened until 1979. This is mainly due to our current patent data source which starts in 1975. This dataset however has the benefit of disambiguated and geographically localized inventors. In future work, we aim at expanding our data set to openings before 1979.

[8]Table B.3 in Appendix B details which sample restrictions account for how many dropped libraries.

all patent libraries and all Federal Depository Libraries in our sample in the continental United States.

Table 2.1 shows summary statistics for patent libraries and matched Federal Depository Libraries in the year before the opening of the patent library. The number of patents per capita, of young, and of small firms are very similar prior to the opening of the patent library. The only difference is that within 25km of the library, the population size is significantly larger for control than for patent libraries, with 1.8 million on average. The reason is that every large city has a Federal Depository Library and thus New York City is included in this subsample. If we compare the median population sizes, cities with patent libraries are with 654 thousand approximately as large as cities with Federal Depository Libraries with a median population of 689 thousand.

Table 2.1: Summary Statistics in the Year Before Opening

	Patent Libraries	Control Libraries	Diff	P-Value
# Patents/100k	10.03	9.42	-0.61	0.65
# Patents	77.22	95.52	18.30	0.25
# Pat. small firms/100k	5.19	4.44	-0.75	0.31
# Pat. big firms/100k	4.84	4.98	0.14	0.87
# Pat. young firms/100k	3.99	3.04	-0.95	0.11
# Patents old firms/100k	6.04	6.38	0.34	0.73
Population in 100k	8.58	18.33	9.76	0.00
Uni Library	0.68	0.63	-0.05	0.49

Notes: This table shows the averages of the data for patent and control libraries. The last two columns shows differences with the associated significance levels. A firm is defined as young if its first patent was filed less than five years before the opening of the patent library. Otherwise it is old. A firm is defined as small if it has no more than 5 patents before the opening of the patent library. Otherwise it is big. The p-values result from a t-test with unequal variances.

In our empirical analysis, we estimate the effect of opening up a patent library on innovation within 25 km around the new library. We can interpret our estimates causally if in the absence of the opening of the patent library the number of patents per capita around the patent library would have had the same trend as the number of patents around the control libraries. One potential concern about this identification assumption might be that a library applied to become Patent Depository Library because its librarians expected that innovative activities in their region would pick up in the future. In contrast, the librarians who did not apply might have had the expectation that patenting might stagnate in their region.

This is possible but does not seem likely. We are looking at the expansion of the patent library program in 1977 that had the aim to open up a library in every state. This program was motivated by equitable access and thus potentially less endogenous to local economic conditions. We also interviewed several librarians and the reasons to become a patent library seemed idiosyncratic: one librarian argued that status considerations between libraries played a role, another said that the librarian in charge wanted to take part in the seminars in Washington DC, and some cited public service considerations.

To show that the assumptions underlying our identification strategy are reasonable, we conduct several plausibility checks after presenting our main results. First, we show that before the patent library was opened, the number of patents filed per capita was the same around the soon to be designated patent library and the control libraries. This speaks in favor of parallel trends. Second, we find little effect if we assign pseudo treatments to the closest control library. This speaks in favor of the SUTVA assumption. Third, we use a host of different specification for the control group and show that our results are robust.

2.4 Results

The opening of a patent library reduced the costs of searching for prior art for inventors in close proximity. In this section we first estimate the overall effect of the opening of patent libraries on the local number of patents. We also show the heterogeneity of these effects along inventor and library types. Subsequently, we analyze how patent libraries change the structure and content of patents in the same region.

2.4.1 Patent Libraries Increase Local Innovation

We start by investigating whether opening a patent library has any impact on patenting within 25 km around the new library. In Figure 2.3, we plot the yearly average treatment effect on the treated of opening up a patent library on the number of patents around patent libraries and around control libraries over time. The time is measured relative to the opening date of the patent library and by the filing year of the patent.

Figure 2.3: Non-parametric Evidence
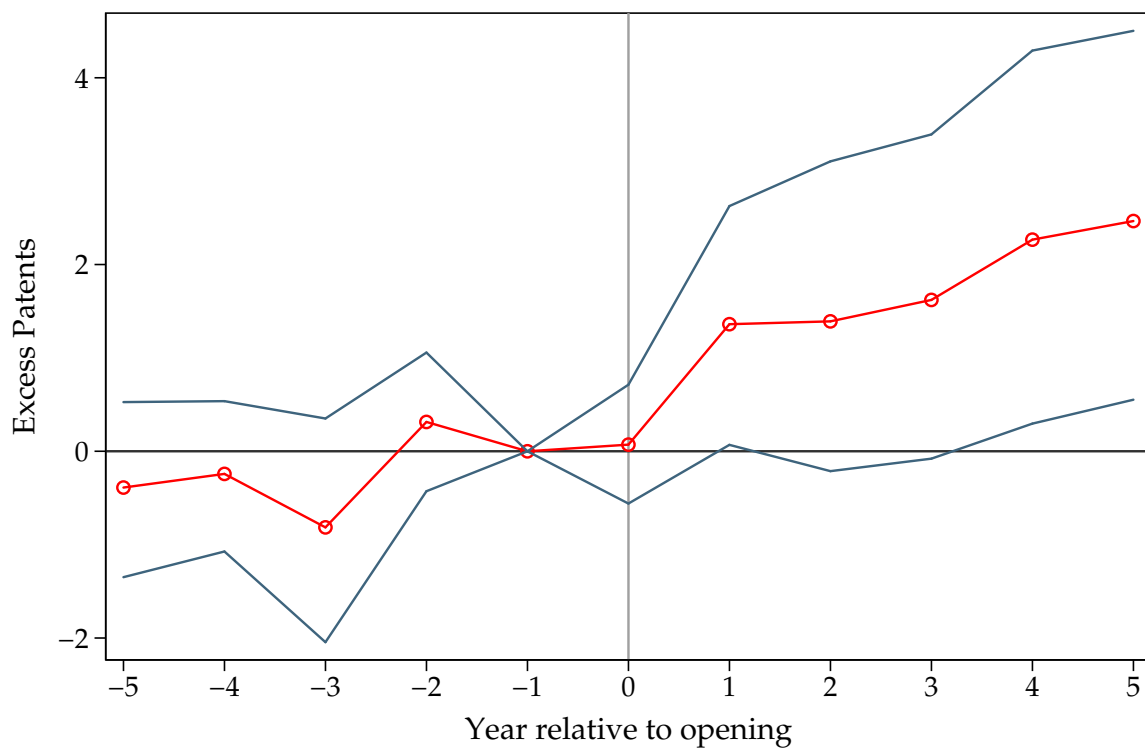
Notes: This figure shows the yearly average treatment effects on the treated of opening up a patent library on the average number of patents within 25 km of patent libraries relative to the average number of patents around matched federal depository libraries. The 90% confidence intervals are based on bootstrapped standard errors. We assign each patent library and all Federal Depository Library within the same state and within 250 km as control group. Data taken from Li et al. (2014). We exclude the patent libraries of Burlington and Rochester.

We find that in the year after the opening, the number of patents around the patent library increases significantly relative to the number of patents around the control libraries. The increase starts in the year after opening and is stable in the following five years. The difference in the number of patents is mostly significantly different from zero on the 10% level. Prior to the opening of the patent library, the number of patents per capita is very similar for treatment and control patents. This speaks in favor of the parallel trends assumption and for a causal interpretation of the estimated effects.

To quantify the size of the effect, we estimate the following difference-in-differences specification where $PatentLib_i$ is an indicator if the library i is a patent library and $Post_t$ is an indicator for the five years after the opening of the patent library:

$$\frac{\#Patents_{it}}{Population} = \alpha_t + \gamma_i + \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + \varepsilon_{it} \tag{2.1}$$

where γ_i are library and α_t are year fixed effects. The coefficient of interest is β_2 and measures the average number of excess patents within 25 km around patent libraries per year in the five years after the patent library was opened. In specifying the standard errors, we allow for clustering at the patent library level (Bertrand et al., 2004).

We report the results for estimating Equation 2.1 in Table 2.2. In column (1) we report our baseline specification where we match to each patent library all Federal Depository Libraries in the same state and within 250 km. We find that on average the number of patents per capita in close vicinity of the patent library increased by 1.8 relative to the control group. This is an increase of over 18% relative to the average patenting around patent libraries before the opening. Around 75% of this increase is driven by small companies (column 2). We define a company as small if it has less than five patents before the opening of the patent library. The effect for large companies in column (3) is smaller and insignificant although they make up more than 50% of all patenting.

Table 2.2: Patent Libraries and Local Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Baseline	Inv. Size		Inv. Age		Inv. Type		University Lib		Patenting	
	250km	Small	Large	Young	Old	Indiv.	Firms	Yes	No	High	Low
Post	0.1	-0.1	0.3	-0.3	0.5*	-0.0	0.1	-0.2	-0.6	0.7	-0.2
	(0.4)	(0.3)	(0.2)	(0.2)	(0.3)	(0.1)	(0.3)	(0.3)	(1.3)	(1.3)	(0.3)
Pat Lib x Post	1.8**	1.1**	0.7	1.0**	0.8	0.3	1.4*	1.9**	1.5	1.6	1.9**
	(0.8)	(0.5)	(0.6)	(0.5)	(0.6)	(0.2)	(0.7)	(0.8)	(2.0)	(2.1)	(0.8)
Constant	6.0***	1.9***	4.1***	-0.3	6.3***	0.8***	5.2***	5.6***	6.3**	13.1***	5.6***
	(0.6)	(0.5)	(0.3)	(0.4)	(0.3)	(0.1)	(0.6)	(0.5)	(2.4)	(1.5)	(0.5)
R2 (within)	0.12	0.19	0.039	0.23	0.056	0.09	0.11	0.14	0.17	0.22	0.14
Clusters	38	38	38	38	38	38	38	26	12	20	26
Obs.	4983	4983	4983	4983	4983	4983	4983	3443	1540	2145	3443

Notes: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is equation (2.1). As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within 250 km as controls. In columns (2) and (3) we split the dependent variable by the size of assignee. An assignee is defined as large if it has more than five patents before the opening of the patent library. In the following two columns we split the dependent variable by young and old assignees. An assignee is young if it filed its first patent less than five years before the opening of the library and old otherwise. In column (6) we only use patents that do not have a corporation as assignee and in column (7) we only use corporate patents. In columns (8) and (9) we split the sample by an indicator if the area of the patent library is also a university library and where it is not. In columns (8) and (9) we split the sample by an indicator if the area of the patent library was above or below the median in terms of patenting five years before its opening. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

If we split the sample by the age of assignee, we find statistically significant effects only for young assignees. We define an assignee as young if its first patent was filed less than 5 years ago. In columns (6) and (7) we separately analyze patents from individual and corporate inventors. The effect is mainly driven by the latter. This might seem puzzling at first, given that we would expect that individual inventors in particular would have problems to access prior art. However, note that young and small *companies* would also qualify as corporate inventors.

In columns (8) to (11), we first split the sample into patent libraries that are associated with a university and into regions which are above and below the median of patents per capita, five years before the opening of the library. The effects are significantly different from zero for university libraries. This is plausible as university students that currently have little access to prior art might have a high potential to innovate. This points to complementarities between access to prior art and technical education. We do not find differential effects for regions with a high or a low patenting rate.

Overall, our results suggest that in particular young and small companies react to the opening of a patent library. This is what we would expect if access to patents increases local innovation. We will return to this when analyzing the structure of patents of new and of incumbent inventors.

Table 2.3: Auxiliary Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Baseline	Patents between	Fake	Matched by		Different control group			All
	250km	25-50 km	Opening	Pop	Patents	500km	State	Within	libraries
Post	0.1	0.3*	0.3	0.3	0.0	0.1	0.3		-0.1
	(0.4)	(0.1)	(0.4)	(0.4)	(0.3)	(0.4)	(0.4)		(0.4)
Pat Lib x Post	1.8**	-0.2	-0.2	1.7**	2.5***	1.8**	1.4	2.3**	4.3
	(0.8)	(0.1)	(0.9)	(0.8)	(0.7)	(0.8)	(0.9)	(1.0)	(3.3)
Constant	6.0***	8.4***	-1.5*	6.7***	5.0***	5.8***	6.0***	7.0***	7.1***
	(0.6)	(0.2)	(0.8)	(0.6)	(0.5)	(0.5)	(0.5)	(2.0)	(0.7)
R2 (within)	0.12	0.04	0.17	0.13	0.14	0.13	0.19	0.38	0.12
Clusters	38	38	36	38	37	39	40	65	40
Obs.	4983	4983	5027	4048	4422	6897	8052	2340	5159

Notes: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is equation (2.1). As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within 250 km as controls. In column (2) we use the number of patents between 25 and 50 km as outcomes. In column (3) we assign a treatment indicator to the FDL closest to the patent library and drop patent libraries from the sample. In column (4) we use FDLs within 250 km and match additionally on similar sized population within 25 km of the treatment and control library. We define a city with a control library as similar if it is less than three times the size of city with the patent library. In column (5) we include only FDLs which had in their close vicinity not more than 50% more filed patents in the year before the opening than patent libraries. In column (6) we extend the control sample to FDLs within 500 km and in column (7) to all FDLs within each state. In column (8) we use not-yet opened patent libraries as control for patent libraries. In column (9) we additionally include Burlington, Vermont and Rochester, NY in our estimation sample. We use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

2.4.2 Plausibility and Robustness Checks Confirm the Results

The main concern for our estimation strategy is that libraries chose to become patent libraries in regions where more innovation was expected in the future. Then, the increase in patenting might have happened independent of the actual opening of the library.

Both the timing and the localization of the increase in patenting make it unlikely that this is the case. First, Figure 2.3 shows that in the five years before the opening of the patent library, there are no systematic differences in patenting between patent libraries and Federal Depository Libraries. In addition, the increase exactly coincides with the opening up of the patent library. This suggests that there are no observeables trends in innovation on which librarians could base their decision and that the expected increase must have exactly coincided with the designation of the patent library. This seems unlikely.

In column (2) of Table 2.3 we also show that the increase is localized in a small area around the patent library. In this regression, we use the number of patents between 25 and 50 km around the treatment and the control library as outcome and do not find any effect. This is also true for wider circles around libraries (Figure 2.4). This implies that the number of patents only increases around the patent library but not in the wider area. As a consequence, if expectations were to explain the increase in patenting, the librarian must have had correct expectation about both, the timing and the area of expected increases in innovation. Again, this seems unlikely.

Another concern might be that we overestimate the effect because opening up a patent library potentially causes inventors to move in space. Then, we might just measure a redistribution of innovative effort in space and not an increase in innovative activity of incumbent potential inventors. Technically, our estimation would then violate the no interference assumption (SUTVA). To show that this is not the case, we drop all patent libraries and assign a fake treatment indicator to the closest Federal Depository Library and re-estimate the effect in column (3). If inventors move towards the library it is reasonable to assume that they would move most from the closest comparable city. Yet, we do not find any effect. This suggests that there is no differential trend between closer control libraries and libraries that are further away and that interference is a second order concern.

Figure 2.4: Treatment Effect by Distance to Library
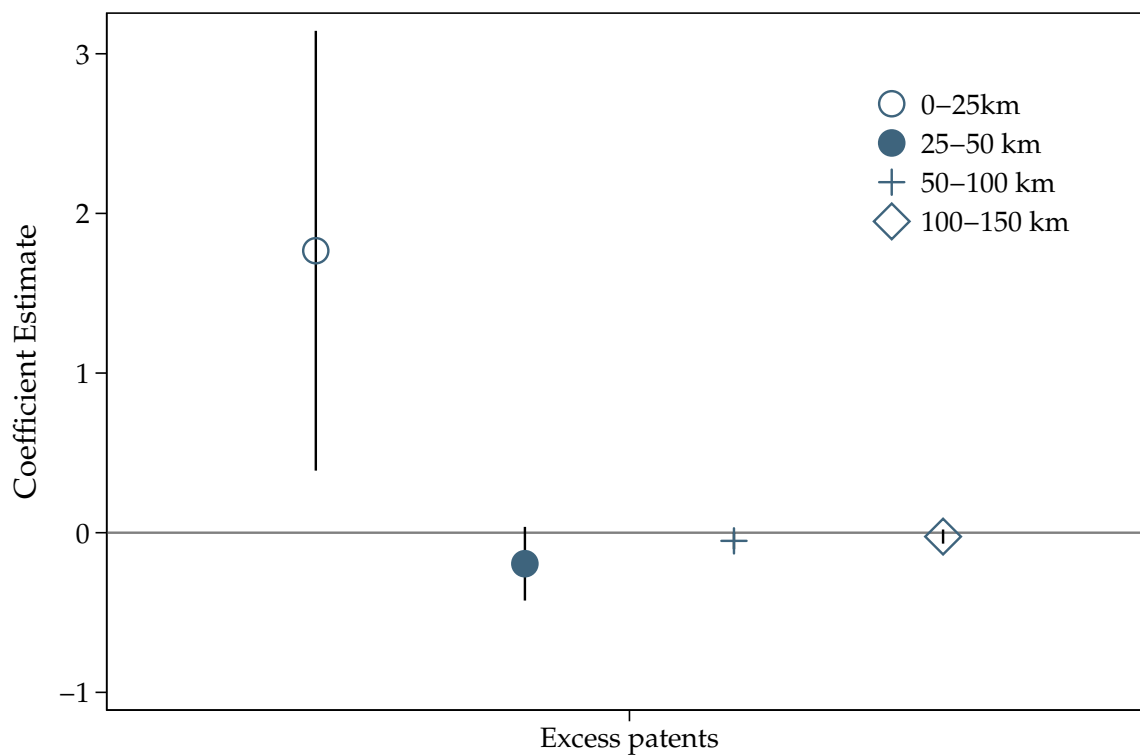
Notes: This figure shows difference-in-differences estimates of opening up a patent library on the average number of patents along different distances to patent libraries relative to the average number of patents in the same distance around matched federal depository libraries. Error bars show 90% confidence intervals with standard errors which allow for clustering at the patent depository library level. We assign each patent library and all Federal Depository Library within the same state and within 250 km as control group. Data taken from Li et al. (2014). We exclude the patent libraries of Burlington and Rochester.

Our estimation strategy is based on the assumption that Federal Depository Libraries in the same state and within 250 km are a suitable control group for patent libraries. In columns (4) to (8) we use different control groups to show that our results are robust. In columns (4) and (5) we use a closer control group. In column (4) we only match Federal Depository Libraries that are in city that is less than three times larger in terms of population than the city of the patent library. In column (5) we only use Federal Depository Libraries that have less than 50% more patents per capita as the patent library in the year before the opening of the patent library. In both cases, the effect is similar. In columns (6) and (7) we use broader control groups. In column (6) we include all Federal Depository Libraries within 500 km and in column (7) all Federal Depository Libraries within the state. The coefficients are in both cases of similar size, yet the estimated coefficient with all control libraries in the state is insignificant. In column (8) we only use other patent libraries as control observations and leverage differential opening dates. The coefficient is around the same size and statistically different from zero.

Lastly, we use a different estimation sample. In our main sample we drop the libraries of Rochester in the state of New York and Burlington in Vermont. The reason is that both have an extremely high patent per capita ratio because they host Kodak, Xerox, and Bausch & Lomb in case of Rochester and IBM in case of Burlington. As a consequence, the control group is too dissimilar. In column (9) we include these two libraries and repeat our main analysis. The effect is much larger but insignificant.

2.5   Patent Libraries Change the Structure of Patents

In this section, we analyze the impact of Patent Depository Libraries on the structure and quality of patents. If improved access to prior art drives our results, we would expect changes along both of these margins. In the following we first show that the geographic distance of cited prior art increases. In a second step, we show that patents around new established patent libraries are cited from patents whose assignees are further away. This suggests that Patent Libraries facilitate access to more distant prior art and increase opportunities for cumulative innovation.

Table 2.4: Impact of Patent Libraries: Backward Citations

Dep. Var.:	\# Citations	Distance		Originality	Cross-Tech
	(1)	(2)	(3)	(4)	(5)
		Full	Small		
Patent library	-1.1***	-159.8***	-118.7***	-2.1**	1.7*
	(0.3)	(35.8)	(31.9)	(1.0)	(0.9)
Post x Patent library	-1.2**	72.2*	95.5**	-0.5	-0.0
	(0.5)	(43.5)	(39.6)	(1.2)	(1.1)
Constant	14.3***	1539.0***	1567.9***	47.1***	26.0***
	(0.0)	(2.7)	(2.3)	(0.1)	(0.1)
Clusters	22843	22665	21681	22843	22843
Obs.	850546	786898	378142	850546	850546

Notes: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. As controls we use fixed effects on the patent library-technology-application year level. In all columns, we use patents by inventors around Federal Depository Libraries (FDLs) within 250 km as controls. Column (1) uses the number of backward citations as the dependent variable. In columns (2) and (3) we use the average distance to the cited patents as the dependent variable. The distance measure is only defined for the subset of citing and cited patents from U.S. inventors. In column (2), we do so for all inventors in the sample. In column (3) we only use small inventors, defined as those assignees who have less than five patents in their portfolio before the library opens. In column (4), we use the mean originality of the patent as defined in Hall et al. (2001) as the dependent variable. In column (5), we use the likelihood of the patents' backward technology class citations relative to its classes' general citation pattern as the dependent variable. This captures "unusual" cross-technology citations. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the level of the fixed effects. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

2.5.1 Patents Cite Geographically More Distant Prior Art

If better access to prior art is driving our results, the type of patents that local innovations build on should change. We would expect local inventors to cite more patents from distant locations and patents from other fields. To test this hypothesis, we construct a sample of all patents around patent libraries and match these to control patents with the same technology class, the same filing year and with inventor located around a Federal Depository Library. As the dependent variable, we use the total number of backward citations, the average geographic distance to the inventor of the cited patent, the originality of the citing patent, and the average number of cross-technology backward citations.[9] We cluster standard errors at the patent library-technology-application year level.

---

[9]The distance measure is only defined for the subset of patents from U.S. inventors. To define originality of a patent, we follow Hall et al. (2001).

Table 2.4 shows the result from a difference-in-differences specification on the patent level. Column (1) shows that the average number of backward citations decreases after a patent library opened by one citation on average. The estimated coefficient is significantly different from zero at the five percent level. Columns (2) and (3) use the geographic distance between citing and cited inventor as dependent variable. Both columns show an increase in the average distance between citing and cited patents, in line with the interpretation that access to previously less available prior art is driving our effects. The effect size corresponds to an increase of around 26% in distance.

Column (3) shows that this effect is again substantially larger for small assignees, reinforcing our interpretation that access to previously unattainable information is driving effects. Columns (4) and (5) show that in terms of originality and non-standard cross-technology citations, there are no significant effects of patent library openings on the structure of patents.

2.5.2 Patents are Cited by Geographically More Distant Inventors

Did opening up a patent library change the quality of new patents in the region? Again, we compare patents with the same technology class and same application year around patent libraries and around a Federal Depository Library before and after the patent library opened. Table 2.5 shows the results of a difference-in-differences specification on the patent level. Column (1) shows that the average number of forward citations does not increase significantly. Column (2) uses the geographic distance of forward citations as the dependent variable. The average distance between patents around depository libraries and the patents which cite them increases, but is only measured imprecisely. The coefficient is not significantly different from zero on the conventional levels but with a p-value of 0.11 is close to the 10% significance level. The effect size corresponds to an increase of around 5% in distance, relative to the baseline. In combination with the results from the previous subsection, this points towards patent libraries fostering the integration of and exchange between geographically distant inventors working in related fields. In line with this, columns (3) and (4) again show that the average generality of a patent and

Table 2.5: Impact of Patent Libraries: Forward Citations

	Forward Citations			
	(1)	(2)	(3)	(4)
Dep. Var.:	# Citations	Distance	Generality	Cross-Tech
Patent library	-2.6**	-150.7***	-2.1***	1.5***
	(1.1)	(44.5)	(0.6)	(0.5)
Post x Patent library	0.2	81.1	1.0	-0.1
	(1.5)	(50.9)	(0.7)	(0.7)
Constant	49.3***	1599.7***	59.7***	19.2***
	(0.1)	(3.2)	(0.0)	(0.0)
Clusters	22846	22741	22846	22846
Obs.	858093	769438	858093	858093

Notes: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. As controls we use fixed effects on the patent library-technology-application year level. In all columns, we use patents by inventors around Federal Depository Libraries (FDLs) within 250 km as controls. Column (1) uses the number of forward citations as the dependent variable. In column (2) we use the average distance to the citing patents as the dependent variable. The distance measure is only defined for the subset of citing patents from U.S. inventors. In column (3), we use the generality of the patent as defined in Hall et al. (2001) as the dependent variable. In column (4), we use the likelihood of the patents' forward technology class citations relative to its classes' general citation pattern as the dependent variable. This captures "unusual" cross-technology citations. In all regressions, we use the weights suggested by Iacus et al. (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the level of the fixed effects. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

the likelihood of non-standard technology class citations do not increase in response to a library opening.[10]

Overall, these results show that in response to patent depository library openings, the geographic distance between patents around the libraries and both their backward and their forward citations increases. This points towards patent libraries facilitating access to distant prior art and increasing opportunities for cumulative innovation.

2.6 Conclusion

The 'grand bargain' in the patent system is that inventors disclose their ideas in exchange for exclusive rights to market their invention for a limited period. Many legal scholars think that disclosure is a significant benefit of the patent system as it helps inventors to avoid duplication and gives them new ideas to recombine with their own. There is however very little evidence whether or not the disclosure resulting from the patent system affects innovation.

This chapter shows that opening up patent libraries increased innovation in the region and helped to disperse innovative activity across the U.S. The results are largely driven by young and small companies, in line with access barriers as the underlying mechanism. We also show that patent libraries were helping co-located inventors to build on prior inventions by improving their access to distant patented knowledge. We thus measure the "enablement effect" resulting from the disclosure of valuable knowledge contained in patents.

Our estimates most likely provide a lower bound for the effect of patents on cumulative innovation through disclosure. First, in many public libraries, the titles and sometimes abstracts of patents were available in technical journals and books. Thus, even without a patent library, there might have been some awareness about inventions made elsewhere. Second, large companies often had their own patent library. Thus we cannot measure the effect of disclosure on the innovation of large businesses.

---

[10]To define the generality of a patent, we follow Hall et al. (2001).

Chapter 3

Labor Mobility and the Productivity of Scientists

3.1 Introduction

Labor mobility of scientists is ubiquitous in science systems around the world. In the production of new knowledge, labor mobility may be especially important in researchers' quest to find new ideas (Kortum, 1997; Weitzman, 1998), complementary co-authors (Wuchty et al., 2007; Jones, 2009; Catalini, forthcoming; Boudreau et al., forthcoming), or a better research environment (Agrawal et al., 2017). However, because of lacking market mechanisms on academic labor markets, it is unclear whether the existing spatial allocation of talent and the current rate of scientist mobility are efficient. As moving is costly, measuring the extent to which labor mobility affects academic productivity is crucial when designing policies for academic labor markets such as the rate and timing of required job mobility. Yet, little is known about the impact labor mobility has on movers' academic productivity.

The econometric challenge when analyzing this question is that scientists self-select into moving. The problem is analogous to general models of self-selection and migration (Roy, 1951; Borjas, 1987; McKenzie et al., 2010): Only those who benefit from moving will incur the substantial personal costs of doing so. Azoulay et al. (2017) study the determinants of scientist mobility using a sample of elite scientists in the life sciences. They find that

movers are more productive than their non-moving counterparts. Academically, mobility is determined by the number and quality of local collaborators, but personal costs are also important. Ganguli (2015a,b) focuses on the international mobility of scientists from Russia after the end of the USSR. She finds that mobile scientists were more likely to be men, young, and more productive than their non-mobile colleagues. This illustrates that comparing movers to non-movers could easily produce spurious impacts of scientist mobility on academic productivity.

In this chapter, we make use of the German system for senior hires at universities to circumvent this problem.[1] German universities always consider at least two candidates for each position. By law, the hiring committee of the university is required to create a ranked list of suitable candidates. Offers are then made in the order of candidates. We use all scientists on this list as counterfactual for the moving scientists. This setup provides two main advantages: On the one hand, it circumvents the problem of selection into moving as all researchers on the list showed interest in moving to the destination university. On the other hand, candidates on the appointment list are qualitatively comparable. The reason for this is that the hiring committee has strong incentives to weed out unsuitable candidates: professors are appointed for life and it is difficult to predict who will accept an offer. Thus, even a low-ranked candidate might receive and accept an offer and stay until retirement. Lastly, hiring no one has an option value because the total number of positions is limited and fixed. Supporting the assumption that researchers are of comparable academic quality at the time of appointment, researchers on the lists in our sample have the same number of citations to pre-move publications in the data on average (Watzinger et al., 2017b).

Our empirical setup uses data on 1'609 lists with 3'850 researchers from one German university between 1950 and 2005. It covers all fields and all appointed professors in this period. In Watzinger et al. (2017b), we link the researchers on these lists with data on their publications between 1965 and 2005. As measure of scientific productivity, in this chapter we use yearly author- and citation-weighted publications of scientists to account for scientific quality and quantity. We then employ an event-study approach

---

[1]The setup and identification strategy of this paper closely follows the companion paper (Watzinger et al., 2017b).

in combination with difference-in-differences estimations to uncover the causal impact of scientist mobility on academic productivity.

We find that after a move, a researcher's productivity as measured by citation-weighted publications increases by around 13% relative to the control group of non-moving scientists. This translates to 0.5 citation-weighted publications per year. The impact seems to increase over time and is statistically significantly different from zero starting around three years after the move. While the average number of citations to publications does not increase for moving researchers, publications in the upper and the bottom third of the citation distribution increase, relative to publications in the medium third. This is in line with an increase in more explorative and less exploitative research. Note that the results reflect estimates of the average treatment effect on the treated (ATT, see Angrist and Pischke, 2008) as all these researchers showed interest in moving to this university. This may explain the rather large estimates. Estimating the ATT of moving on academic productivity is however policy relevant: It gives an estimate of how much less productive a moving scientist would have been had she not been appointed.

Our results uncover important effect heterogeneity with respect to academic field, academic influence, and type of position. The estimates are entirely driven by researchers in the natural sciences, but do not depend on any single field within this group. In other fields, we do not find any effects of scientist mobility on productivity. With respect to the scientific influence of researchers ("quality"), the increase in academic productivity completely stems from researchers from lists with an above-average number of citations to pre-move work. The effect is driven by researchers who applied for non-chaired positions, and is neither driven by publications where the researcher is the last-named author, nor by publications where she is the first-named author. This suggests that an alternative explanation of the effects as returns to lab ownership is not driving the results.

We provide evidence that the identification assumption of parallel trends in productivity between movers and non-movers in the absence of the move holds. First, we make use of our event-study approach and find that the estimated impact of moving on academic publications is not significantly different from zero in the pre-move period. The impact of moving on academic productivity is only significantly different from zero starting around

three years after the move. Second, when assigning the treatment dummy to the highest ranked non-mover, we do not find any effect of mobility on subsequent productivity. This speaks against an interpretation where higher rank simply reflects higher expected future productivity.

The main alternative channel through which these results could emerge is larger access to capital generated through negotiations with the new university. While we cannot exclude this factor as the driving force of our results, we do not think this explanation is likely: In the subsample of lists where the highest ranked researcher did not accept the offer and a lower ranked researcher moved to the university, this alternative interpretation would predict that the effect size should be smaller. After all, researchers who decline mostly do so after negotiations with their home (or other) universities which increases access to capital. Our results show that the effect is actually *larger* in this subsample. This result also shows that our estimates are not driven by star researchers moving to this university, further corroborating the parallel trends assumption. Finally, using publicly available partial data on lists of rejected offers, we also show that the effect is not restricted to this one university.

The small literature trying to assess the impact of scientists' mobility on their academic productivity has been struggling to find plausible identification strategies and data which meets the stringent requirements to answer this question. This is especially true in light of findings in the migration literature that movers are different from non-movers not only along observable, but also along unobservable characteristics (McKenzie et al., 2010). Hoisl (2007) finds that an inventor's move increases her productivity. She uses the regional characteristics of the invention as an instrument for mobility. However, the exclusion restriction seems unlikely to hold as the selection into patenting likely differs between rural and urban areas.[2] Franzoni et al. (2014) use a survey of active researchers and

---

[2]Trajtenberg (2005) finds that moves increase the value of patents at the new place of the inventor. He reports that movers are selected on a number of characteristics, such as the importance and originality of their inventions. Hoisl (2009) uses survey data on German inventors and finds that after an inventor moved, her number and quality of patents increases relative to a control group of non-moving researchers. This control group matches inventors on their age, their educational background, and their main technical area. Because movements are endogenous, this still leaves the possibility that only those researchers who benefit from the move are mobile. Baeker (2015) studies the mobility of post-doctoral researchers in Germany. Using matching techniques, she finds that changing affiliation decreases the productivity of researchers in the short run.

instrument (international) researcher mobility by having migrated as a child. They find that international mobility increases academic productivity. However, the validity of the exclusion restriction is unclear. It is also unclear how to interpret these local average treatment effects. Finally, Ganguli (2015a) makes use of the collapse of the Soviet Union to assess whether international labor mobility of scientists affects their productivity. She employs individual fixed effects combined with difference-in-differences regressions and finds that after a move, scientists produce around 0.5 more publications per year. In her setup, movers and non-movers differ on a variety of observable characteristics, a fact that does not disappear even in coarsened matched samples. This caveat, in combination with the manifold changes in the former Soviet Union after its collapse, make it unclear how the results translate to labor mobility of scientists more generally. Therefore, there is still little rigorous evidence on the productivity impacts of scientist mobility.

This chapter contributes to this literature in three important ways. First and most importantly, the German university system provides a credible setup to circumvent inherent identification problems. All researchers on the appointment lists specifically applied for the position. This accounts for the problem that scientists self-select into moving. Because of the inherent incentives in the hiring process, researchers on the appointment lists are furthermore similar in academic quality. However, this setup comes at the cost of providing estimates among a group with likely large average treatment effects, relative to the population of scientists. This is mitigated by the German university system which provides ample incentives to move. The selection of movers in German academia is therefore not substantially different from the general population of scientists. Also, if runners-up move to other universities in response to not being appointed at this university, this should work against finding any effect (similar to *substitution bias* in the program evaluation literature, see e.g. Heckman et al., 2000). In summary, runners-up provide a credible counterfactual for movers with unclear implications for scientists who did not show interest in the positions.

Second, this is one of the first analyses to assess the heterogeneity of impacts of labor mobility on innovation across different fields and across differentially influential researchers. While scientist mobility seems to be important for the natural sciences, we find no effect for other fields. The impact of mobility on productivity is driven by researchers from

Figure 3.1: Procedure for Appointing a Professor in Germany
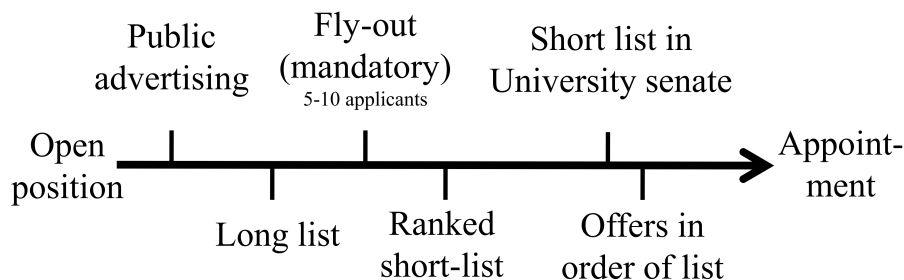
appointment lists with above-median academic impact. This stands in contrast with findings by Ganguli (2015a). Because the literature has so far focused on patenting inventors who mostly work in the natural sciences, the heterogeneity of estimates provides new insights into which scientists benefit from labor mobility.

Third, this is one of the first studies to rigorously estimate the impact of labor mobility for the productivity of *academic* researchers. Research on the impact of mobility on inventive activity has hitherto mostly focused on patenting inventors. However, because academic labor markets mostly lack market mechanisms, the labor mobility of scientists is especially prone to inefficiencies in the spatial allocation of researchers. While most science systems in the world incorporate mechanisms to increase academic mobility, little is known about the impact this has on the movers' scientific productivity.

The remainder of this chapter proceeds as follows. In the next section, we present the institutional setup and the dataset. In section 3.3, we discuss the identification strategy. In section 3.4, we present results, robustness checks, and evidence for the validity of the identification assumption. Section 3.5 concludes.

3.2 The German System for Hiring Researchers Provides a Natural Experiment

In Germany, almost all university professors are civil servants and thus are hired in a highly regulated multi-step process (Figure 3.1).[3] The procedure is designed to give every qualified applicant equal access to jobs in the public service independent of personal con-

---

[3]This section is adjusted from the companion paper, (Watzinger et al., 2017b).

nections. To implement equal access, every open position must be advertised in a national newspaper. The advertisement must contain a list of criteria by which the candidates are compared in the remainder of the process. These criteria for example usually include publications in refereed journals or experience in raising third-party funding. Using these criteria, the hiring committee creates a long list of five to ten candidates who are invited for fly-outs. After the fly-outs, the hiring committee creates a ranked short-list of two to four candidates. All candidates on the short-list are reviewed by at least two external referees who suggest their own independent ranking of the candidates. The ranked short list and the reference letters are submitted to the university senate for review. If the ranked short list is approved by the senate, offers are made to the candidates on the short-list in order of the rank. The first candidate to accept is then appointed.

The hiring process contains several mechanisms to make the process objective and fair. First, internal candidates are usually not eligible to apply for tenured positions, so almost all new professors are external hires from other university. Internal promotions are in theory possible, but must follow much more stringent rules than external hires.[4] Second, the composition of the hiring committee is fixed and contains external members. The hiring committee has at least one professor in the same field but from another university, one member of the university senate from another field, a women's representative, a representative of non-tenured scientific employees, and one undergraduate student representative. Third, the whole process is subject to court review: if one of the non-appointed candidates suspects that the university did not follow due process, the candidate can sue for non-appointment of the chosen candidate, compensation and invalidation of the list ("Konkurrentenklage").

According to the *Gemeinsame Wissenschaftskonferenz*, a joint institute of all German universities and the German government, every open professorship attracted on average 41.8 applications in Germany in 2013 (Wissenschaftskonferenz, 2014). Around 10% of these candidates were considered suitable for the short list, which implies that the average list had four candidates. Of all candidates, 45% received an offer for the position at one

---

[4]The rules are so stringent that for all practical purposes this is perceived (and labeled) as forbidden (Ban of internal promotions - "Hausberufungsverbot"). This rule prevents nepotism because it requires researchers to move at least once in their academic career to get a tenured position. In this chapter, we exclude all internal promotions.

point in time. If a candidate received an offer, the probability that she accepted was around 50%.[5]

Other Researchers on Appointment Lists Serve as Counterfactuals for the Mover

Due to the incentives inherent in the legal set-up of academic appointments in Germany the candidates on the short-list are not only acceptable, but also similar in academic "quality". This provides the opportunity to use all other candidates on the list as a counterfactual for the moving candidate.

The main reason is that the hiring committee has strong incentives to weed out unsuitable candidates. In Germany, all professors are appointed for life and for the hiring committee it is very difficult to predict who will accept an offer. Candidates can receive competing offers during the selection process and thus might have received better offers once the process is complete. What is more, receiving an offer from a different university opens the door to renegotiation at the current university. Even a low-ranked candidate might therefore receive and accept an offer and stay until retirement, as appointment to civil service is for life. As a consequence, the hiring committee has incentives to only put acceptable candidates on the list. Even more, hiring no one has an option value because the number of professorships is restricted. Therefore, if there is a suitable candidate in the following year, the university might not be able to hire her, because all positions are filled. As a consequence, the hiring committee might choose to hire no one.

A second reason why researchers on the list are similar is that in all but rare cases, only candidates who applied to the position are put on the list and the possible salary and teaching obligations are fixed in a pre-determined range. Only researchers who are interested to work at this university for the offered conditions thus apply. Each position in Germany is associated with a category determining the possible salary range, the

---

[5]The data is not exact as open professorships are counted in a two year interval. According to the data 1'612 professors were appointed in Germany in 2013. The universities received 67'117 applications for these positions. 6'954 researchers made it on the ranked list and 3'175 received an offer.

pension benefits and the status of a professor.[6] Even more, at the time of this study, there was salary cap: No professor could earn more than category B10 which corresponds to the category of a General or an Admiral in the German military. Taken together, these restrictions lead to a self-selection of applicants ensuring a more homogeneous applicant pool.

Movers are Similar to the General Population of Researchers

In the traditional German university system, there is no internal promotion to tenure. The only way to receive a tenured position is to accept an offer from another university. As a result, all junior researchers in Germany move between universities at least once. Also, a senior researcher who wants a higher salary or more access to research funds must have at least one outside offer to be able to renegotiate her current contract. As universities often do not renegotiate or the hiring university has more funding, many outside offers are taken. Many researchers move three or four times in their career as a result. Note that if the runners-up in our data also move to other places after not being appointed at this university, this works against finding any effect of moving on subsequent productivity.

These legal restrictions imply that the selection of movers is not very different from non-movers as moves are the norm rather than the exception. Because in other settings, moving researchers differ markedly from non-movers, our setup provides a clear advantage over existing research (e.g., Ganguli, 2015a).

Data

For this study, we have access to all short lists of candidates of one German university from 1950 to 2005. The university under consideration offers a wide range of subjects from

---

[6]For example, the highest remunerated professorship today is the W3 professor which is comparable to a tenured full professorship at a public doctoral institutions in the U.S. A W3 professor is usually expected to lead a research group. The associated salary usually ranges between Euro 60'000 and Euro 110'000 depending on the federal state and seniority level. An overview of salaries for different salary category can be found here: http://www.w-besoldung.net/.

humanities to natural sciences and medicine. In total we have access to 1'609 ranked lists for professorships containing 3'850 researchers (2.4 per list). We match all researchers on the lists with their academic publications from Scopus and the year of their move to the university from historical course catalogs.[7] In our data, over 80% of publications are journal articles. Around 6% of publications are conference proceedings, another 7% are reviews or surveys, and around 1% of publications are books or book chapters. We do not discriminate between different publication formats as these may reflect different publishing standards across fields. The main dependent variable is the author- and citation-weighted yearly number of publications.[8] For all publications, we therefore collect all citations by other publications. As citations are at the moment only available from 1980s onwards, this leaves us with 1'012 short lists with 2'760 scientists.

However, there is a number of observations which cannot be used for inference. We delete lists on which there are no movers or on which there is only one person (leaving 2'218 researchers). We also delete all lists with researchers who did not have registered publications in Scopus in the ten years before or after the move because this indicates either different publication standards in a field or (more likely) matching errors between the Scopus data and the names on the appointment lists (leaving 1'416 researchers). Subsequently, we delete those lists which include scientists who are assigned to more than one appointment list (leaving 1'176 researchers). We also refrain from using eight lists where the discrepancy between mover and non-mover productivity is very large before the move, five researchers who are outliers in terms of pre-move citations, and 6 lists which are incomplete (leaving 1'141 researchers).[9] Finally, we also delete lists with a "Sperrvermerk" which means that the list is sent back to the department after a certain candidate declined.

These data restrictions leave 1'000 researchers on 317 lists in the sample (3.15 per list). While the restrictions are stringent and leave not even half of the original sample in the data, the strong identification assumption detailed below makes a rigorous treatment of

---

[7]More details on the matching process and the construction of the dataset are provided in Appendix C.1.

[8]Author-weighting means that if there are N authors who jointly publish a publication in any year, each researcher gets assigned $1/N$ publications in this year.

[9]Results are similar when not excluding these lists and researchers.

Table 3.1: Descriptive Statistics - Author Level

	Mean	SD	Min	Max
Ln(Pub.+1)	0.60	0.42	0.00	2.16
Ln(Wgt. Pub.+1)	2.08	1.46	0.00	5.94
Ln(Cit. per paper+1)	2.68	1.09	0.00	5.34
No. Authors	1.52	0.51	0.69	5.71
Science	0.68	0.47	0.00	1.00
High Qual.	0.54	0.50	0.00	1.00
Mover	0.36	0.48	0.00	1.00
Observations	1000			

Notes: This table shows descriptive statistics on the author-level for researchers in our sample. Publications are author-weighted. Weighted publications are author- and citation-weighted. The values for all time-varying variables are averages across the five years before and five years after the move. The data stems from all appointment lists of one large university in Germany. Publication data stems from Scopus.

the data necessary.[10]

Because Scopus provides comprehensive data on scientific publications covering all main outlets such as articles, books, conference proceedings, and handbook articles, we infer from missing author-year combinations in the publication data that the author has not produced a scientific publication in this year. We therefore extend our data to a balanced sample covering the 1'000 researchers in our sample over ten years before and ten years after the move.

The author-level descriptive statistics of the sample are shown in Table 3.1. The summary statistics relate to the five years before and after the move as this will be the sample we use for our difference-in-differences strategy. The average researcher in the sample has around 0.6 yearly author-weighted publications, around 4.7 yearly author- and citation-weighted publications and produces publications which are cited around 6.3 times on average. The average number of authors per publication is 5.6. Two-thirds of the sample are researchers in the natural sciences, slightly more than half of all researchers are on lists with above-median pre-move impact as measured by forward citations, and around 36% of researchers in the final sample moved to the destination university.

To assess whether our results are restricted to this university, we also use data from the

---

[10]There are some professors who are appointed from other universities and institutes close to this destination university. When deleting these lists from the sample, results are qualitatively robust (not shown).

magazine "Forschung und Lehre" which regularly reports appointment offers and whether they were accepted or not. We make use of repeated reports on the same position (rejected offers) to construct partial appointment lists. After deleting eight lists which are outliers, we are left with 147 lists providing information on 340 scientists in this sample.

3.3 Empirical Strategy

The main identification problem is that researchers do not move randomly. In particular, because of the heavy costs of moving (Azoulay et al., 2017), only those who benefit from the move will decide to incur these. For example, these might be researchers whose research focus is a superior fit at the destination university. Also, only scientists who are particularly productive researchers might be offered positions. These factors likely lead to overestimation of productivity effects when only comparing moving to non-moving scientists.

To answer the question whether scientist mobility affects academic productivity, an ideal experiment would randomly make some researchers move while keeping others at their current affiliation. In this case, a simple comparison of the academic productivity between movers and non-movers would return the causal effect of scientist mobility on academic productivity. While this ideal is impossible to reach, the setup in this chapter circumvents many of the usual problems when answering this question and allows estimates which are directly policy relevant.

We use runners-up for the position of the appointed professor to construct a close control group for the moving researcher. All candidate researchers applied for the position, were extensively vetted, and were accepted by the university for the short list. Therefore, all of them might have been appointed. Thus, we are able to control for the endogenous selection of researchers to jobs as the runners-up provide a counterfactual for how much the moving researcher would have published in the absence of the move.

In the main specification, we use an event study design combined with a difference-in-differences strategy with the moving professors as treatment group and the researchers who are runners-up as control group. The event study framework has the advantage that

it permits visual inspection of pre-existing trends. Because at points in time further away from the move, the likelihood of confounding factors increases, we only use the five years before and after the move for our preferred estimates. For the annual treatment effects, we use the ten years before and after the move to allow for closer examination of pre-trends. As dependent variable, we use scientific publications of movers and non-movers to measure academic productivity. To adjust for differences in productivity which are solely due to larger research teams in some fields or different collaboration patterns, all publication counts are weighted by the number of authors on a publication.

Publication counts do not account for the underlying quality of research. The setup of hiring researchers in Germany leads to appointment lists with researchers who are similar in academic quality, not necessarily in the absolute number of publications written. Therefore, our preferred measure of academic productivity is a quality-adjusted publication count. To measure quality, we weight publications by the number of their subsequent citations. Using alternative measures such as journal rankings is difficult because across different fields, the main format of academic publications differ. While there certainly are differences in citation standards between fields, citation-weighted publications are still more interpretable and reflect academic publications weighted by their subsequent knowledge flows to other researchers. What is more, this is the preferred measure in the literature using academic publications as the outcome of interest (e.g., Stephan, 1996). Because publications, especially quality-weighted publications, and citations are highly skewed in the data, we take natural logarithms of all dependent variables adding one to all observations to account for years in which researchers did not publish.[11]

The validity of the approach depends on the assumption that in the absence of the move, the (quality-weighted) publications of movers and non-movers would have followed the same trend. Note that the identification condition in this chapter therefore differs from Watzinger et al. (2017b). In Watzinger et al. (2017b), the validity of the approach hinges on the assumption that in the absence of the move, citations to the pre-move work of all ranked candidates would have followed the same trend. Here, our assumption does

---

[11]This non-linear transformation of the dependent variable may in principle lead to biased estimates (e.g., Santos Silva and Tenreyro, 2006). Results are robust to estimating the specification in levels and to using alternative estimation methods for count data such as fixed effects Poisson regressions (Hausman et al., 1984; Wooldridge, 1999; Azoulay et al., 2010). See Table C.1 in Appendix C.2.

not only relate to the quality of pre-move publications, but encompasses that researchers would have followed the same *future* academic productivity in the absence of the move and is therefore substantially stronger.

This assumption is by definition untestable. In the latter part of the chapter, we provide a series of robustness checks to show that the assumption is plausible. For example, we show that all non-movers have the same trend and thus rank does not predict increases in productivity per se. What is more, we apply more stringent restrictions on our dataset than in Watzinger et al. (2017b) to ensure that candidates are not only similar with respect to the citation potential to their pre-move work, but also similar in terms of their expected future productivity.

3.4 Results: Moving Researchers Become More Productive

Labor mobility may be important for the productivity of researchers because it could allow access to new ideas (Kortum, 1997; Weitzman, 1998), to complementary co-authors (Jones, 2009; Catalini, forthcoming), or to a better research environment (Agrawal et al., 2017). In this section, we estimate whether scientist mobility indeed increases academic productivity. To measure academic productivity, we rely on researchers' yearly publication output, the main goal of academic research. We adjust publications for the number of authors in all specifications and for their future citation count in our preferred estimates.

3.4.1 Publications Increase in Response to the Move

In this subsection, we use regression models with time-varying coefficients to estimate the impact of moving on the number and quality of academic publications, the main output of academic research.

To analyze the change in academic productivity in response to the move, we estimate the following econometric model:

$$Log(Publications_{i,t,l} + 1) = \alpha \cdot Move_i + \beta_{t \neq 0} \cdot Move_i + \delta_t + \gamma_l + \epsilon_{i,t,l} \qquad (3.1)$$

where the dependent variable $Log(Publications_{i,t,l} + 1)$ is the natural logarithm of either the author-weighted or the author-and-citation-weighted number of publications of researcher i in year t (relative to the move) on list l. This specification therefore uses an event-study approach, with the move being the event of interest. The coefficients of interest are the β_t which capture the average yearly difference in the dependent variable between the mover and the other researchers on the same appointment list.[12] The specification controls for list fixed effects γ_l, such that all differences across lists which are constant over time are differenced out. Therefore, this specification essentially compares movers and non-movers on the *same* appointment list. We also control for period fixed effects δ_t which control for changes in productivity around the appointment which are common to all researchers in our sample. Finally, the error terms $\epsilon_{i,t,l}$ allow for clustering at the appointment list level, which in this setup nests adjustments for clustering at the researcher level and therefore produces rather conservative estimates of the standard errors (Bertrand et al., 2004).

In Figure 3.2, we show estimates for all β_t where the dependent variable is the number of author-weighted publications of a researcher. Prior to the move, the publications of movers and non-movers are not systematically different in any single year. After the move, the number of publications of the movers increases significantly relative to the control group. The effect becomes significant in year three after the move and is increasing over time. With respect to the timing of the impact, one should keep in mind that the effect is primarily driven by researchers in the natural sciences where the publication process is substantially faster than in other fields, such as economics (cf. Stephan, 1996). Therefore, it is unlikely that this effect merely reflects a strong "pipeline" of pre-move research results which led to the appointment in the first place.

Because raw publication numbers do not reveal anything about the underlying quality of publications, in Figure 3.3 we use author- and citation-weighted publications as the dependent variable. The coefficients reveal a similar pattern: While the number

---

[12]We use the year of appointment as the baseline year. Results are robust to using other periods as reference.

Figure 3.2: Excess Publications of Movers
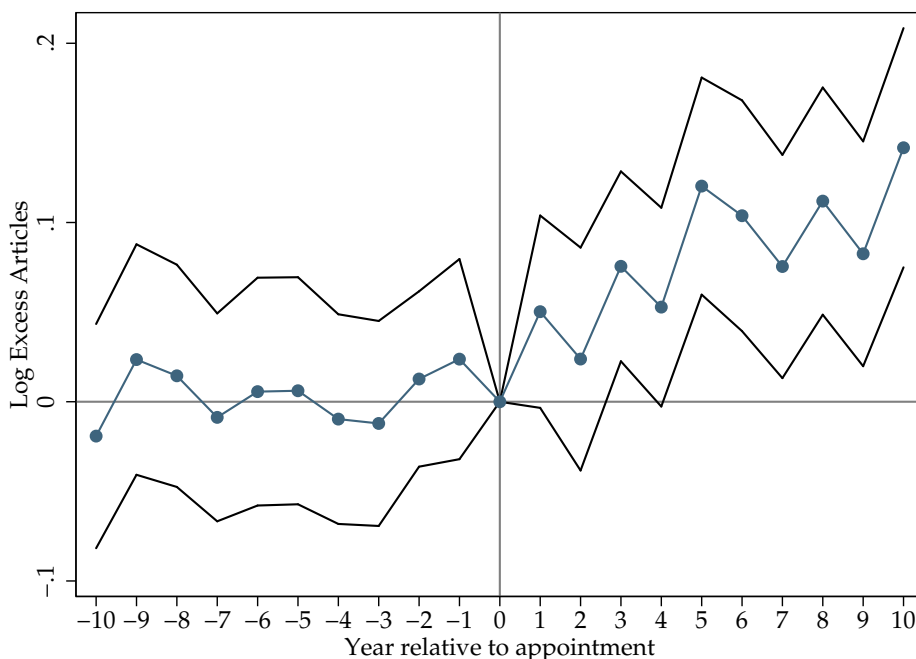
Notes: This graph shows the estimated number of yearly log excess publications by moving researchers, relative to the year of the move. Publications are weighted by the number of authors. To arrive at these estimates, we regress the yearly log number (plus one) of author-weighted publications by researchers on the appointment lists on an indicator variable equal to one if the researcher moved (interacted with time dummies), year fixed effects, and appointment list fixed effects. The dark blue line represents the 90% confidence bands for the estimated coefficients. Standard errors allow for clustering at the appointment list level. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

of citation-weighted publications does not significantly differ between movers and non-movers before the move, it is higher for movers in response to the move. The coefficient estimates are consistently significantly different from zero from year four after the move. Taken together, these estimates suggest that mobility of scientists does increase academic productivity.

Publications Increase by 13% in Response to the Move

The approach above provides graphical evidence that mobility increased academic productivity, but has rather low statistical power due to the estimation of year-by-year coefficients. To quantify the absolute size of the impact shown in the previous subsection, we therefore estimate the average change in productivity per year in response to a re-

Figure 3.3: Citation-weighted Excess Publications

Notes: This graph shows the estimated number of yearly log excess publications by moving researchers, relative to the year of the move. Publications are weighted by the number of authors and by the number of follow-on citations. To arrive at these estimates, we regress the yearly log number (plus one) of author- and citation-weighted publications by researchers on the appointment lists on an indicator variable equal to one if the researcher moved (interacted with time dummies), year fixed effects, and appointment list fixed effects. The dark blue line represents the 90% confidence bands for the estimated coefficients. Standard errors allow for clustering at the appointment list level. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

searcher's move in a difference-in-differences model. As pre- and post-periods, we use the five years before and after the move. We estimate the following specification:

$$Log(Publications_{i,t,l} + 1) = \alpha \cdot Move_i + \delta \cdot Post_t + \beta \cdot Move_i \cdot Post_t + \gamma_l + \epsilon_{i,t,l} \quad (3.2)$$

where $Move_i$ indicates whether the researcher at hand moved and $Post_t$ is an indicator for the time period one to five after the move. Again, $Log(Publications_{i,t,l} + 1)$ refers to the natural logarithm of author- (and citation-) weighted publications of researcher i on list l in year t. The coefficient of interest is β, the average difference in post-move log publications between movers and non-movers. As controls, we use list fixed effects

γ_l, therefore only identifying from within-list changes in academic productivity after the move. Note that the coefficient estimates for β in this specification are equivalent to an estimation with individual fixed effects because the mover dummy accounts for mean differences between movers and non-movers within lists.[13] However, estimating this mean difference explicitly allows for testing whether the difference in academic productivity between movers and non-movers before the move is similar. These estimates shed light on the plausibility of the identification assumption. Again, the error term $\epsilon_{i,t,l}$ allows for clustering at the appointment list level.

Table 3.2: Excess Publications of Moving Scientists

Dep. Var.:	Ln(Pub.+1)	Ln(Citation-Weighted Publications+1)				
	(1)	(2)	(3)	(4)	(5)	(6)
Mover	0.04*	0.07	0.11	-0.01	0.11	0.03
	(0.02)	(0.07)	(0.09)	(0.10)	(0.12)	(0.07)
Post	0.00	-0.03	-0.10*	0.12**	-0.16**	0.10**
	(0.01)	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)
Post x Move	0.06***	0.19***	0.26***	0.04	0.32***	0.06
	(0.02)	(0.06)	(0.07)	(0.08)	(0.09)	(0.07)
Constant	0.53***	1.87***	2.32***	0.98***	2.79***	1.00***
	(0.01)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)
List FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Baseline	Baseline	Science	Humanities	High Qual.	Low Qual.
Adj R2	0.34	0.45	0.36	0.40	0.24	0.33
Lists	317	317	214	103	156	161
Observations	10000	10000	6780	3220	5350	4650

Notes: This table shows fixed effects regressions on the author-by-year level. The data comprises the five years before and five years after the appointment. The dependent variable is the yearly log number (plus one) of author- (column 1) and author-and citation-weighted (columns 2-6) publications of authors on the appointment list. In columns (3) and (4), we split the sample into researchers in the natural sciences and all other fields. The natural sciences are comprised of Biology, Chemistry and Pharmacy, Medicine, Physics, and Veterinary Medicine. In columns (5) and (6), we split the sample along the median number of listwise average citations to pre-move publications of researchers on this appointment list. Move is an indicator if the researcher was appointed, Post is an indicator for the post-move period and the indicator of interest PostxMove is the interaction for the period after the move of treated authors. In every regression we control for fixed effects for the appointment list under consideration. The constant reflects the intercept that makes the prediction calculated at the means of the independent variables equal to the mean of the dependent variable. Standard errors allowing for clustering on the appointment list level are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

---

[13]As such, controlling for (individual) covariates would not change the coefficient estimates for β but would merely affect the standard errors. This implies that our estimates are robust to the inclusion of list rank fixed effects or origin university fixed effects, for example. Results from these specifications are available on request.

The results of this specification are reported in Table 3.2, where the dependent variable is the author-adjusted number of publications. Column (1) shows that after the move, movers have 0.06 more log publications on average than non-movers. In comparison to the baseline publication rate, this is an increase of around 30%. This seems very large, but translates to around 0.1 more author-weighted publications per year. These results are smaller, but similar to Ganguli (2015a), who finds around 0.2 more publications per year for scientists who left the Soviet Union relative to those who stayed but could collaborate with foreign researchers. Column (1) also reveals that movers have slightly more absolute publications than non-movers prior to the move, an effect that is statistically significant at the ten percent level. Column (2) shows the specification using the natural logarithm of author- and citation-weighted publications as the dependent variable. Again, movers are significantly more productive after the move. The effect is smaller, translating in around 13% more citation-weighted publications per year relative to the baseline effect. While this effect may still seem large, note that it reflects the ATT of moving among researchers applying for a job at a different university (similar in spirit to leveraging oversubscribed lotteries). Because researchers move for a reason, the ATT is likely larger than the average treatment effect for scientists who are not interested in moving to the university. The column also shows that the difference in the baseline rates of quality-weighted publications between movers and non-movers is not significantly different from zero. This is consistent with the identification assumption of parallel trends between movers and non-movers in the absence of the move.

To assess the origin of these large effects, we analyze the impact of moving across fields and academic impact. The results of this exercise are reported in columns (3) through (6). Column (3) shows that the effect is entirely driven by the natural sciences, whereas the impact of moving is insignificantly different from zero in other fields. This can also be clearly seen in Figure 3.4, where we show time-varying coefficients as described before for the natural sciences and for other fields. For the natural sciences, the impact is significantly different from zero from year three after the move and is less noisy than in the baseline estimates. In contrast, the estimated treatment effects for all other fields are never significantly different from zero.

In columns (5) and (6), we split the sample along the median number of citations (by list)

to pre-move work. The effect is entirely driven by and only statistically significant for researchers on lists with high average academic influence before the move, measured as being above median in the forward citation distribution. Therefore, especially high-impact researchers seem to benefit from moving (to this university). This stands in contrast to findings by Ganguli (2015a), who shows that among scientists leaving the Soviet Union after its collapse, the move decreased the productivity of highly active researchers.

In summary, these results show that the productivity of moving researchers increases in response to the move, relative to the control group of non-moving researchers on the same appointment list. These effects are driven by the natural sciences and by researchers on lists with above-average academic impact.

3.4.2 Quality and Collaboration

In this subsection, we analyze how moving impacts academic productivity along the quality distribution of publications and whether moving increases the size of research teams. Column (1) of Table 3.3 shows the baseline estimates using citation- and author-weighted publications as the dependent variable. The second column uses the average citations to publications of active researchers and shows that these did not increase. This implies that the average quality of articles did not increase.

However, note that more novel research displays higher variance in their citations (Wang et al., 2017). That is, while they are more likely to be highly cited, they are also more likely to receive a very low number of citations. Columns (3) to (5) therefore use publications in the top third, medium third, and bottom third as the dependent variables, respectively. In line with the idea that moving generates access to new ideas and increases the novelty of new publications, these columns show that the impact of moving is statistically significant for publications in the top and the bottom third of the citation distribution of articles of researchers in our sample. The effect is smaller and only significantly different from zero at the ten percent level for the medium third of publications. This is in line with the idea that the increase in quality-weighted publications is driven by access to new ideas.

Finally, in column (6), we analyze whether the effect is driven by the average number of

Table 3.3: Excess Publications of Moving Scientists: Heterogeneity of Effects along Publication Quality and on Team Size

Dep. Var.:	Ln (Cit.-Wgt.+1)	Ln (Cit.+1)	Ln(1+Pub. with Citations in..)			Ln(1+ #Authors)
			Top	Medium Third	Bottom	
	(1)	(2)	(3)	(4)	(5)	(6)
Mover	0.07	0.02	0.01	0.02	0.03**	0.01
	(0.07)	(0.06)	(0.01)	(0.01)	(0.02)	(0.02)
Post	-0.03	0.01	0.01	0.01*	-0.00	0.11***
	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Post x Move	0.19***	0.04	0.03**	0.02*	0.04***	0.04
	(0.06)	(0.05)	(0.01)	(0.01)	(0.02)	(0.02)
Constant	1.87***	2.77***	0.23***	0.21***	0.19***	1.51***
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
List FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Baseline	Active	Baseline	Baseline	Baseline	Active
Adj R2	0.45	0.45	0.38	0.22	0.25	0.54
Lists	317	312	317	317	317	312
Observations	10000	7263	10000	10000	10000	7263

Notes: This table shows fixed effects regressions on the author-by-year level. The data comprises the five years before and five years after the appointment. The dependent variable the natural logarithm (plus one) of author- and citation-weighted publications in the first column. It is the natural logarithm (plus one) of the average number of follow-on citations in column (2). The dependent variable is the natural logarithm (plus one) of publications with citations in the upper third (column 3), medium third (column 4), and bottom third (column 5) of the citation distribution of all publications of researchers in the sample in the following columns, and the natural logarithm (plus one) of the average number of authors on a publication in column (6). Move is an indicator if the researcher was appointed, Post is an indicator for the post-move period and the indicator of interest PostxMove is the interaction for the period after the move of treated authors. In every regression we control for fixed effects for the appointment list under consideration. The constant reflects the intercept that makes the prediction calculated at the means of the independent variables equal to the mean of the dependent variable. Standard errors allowing for clustering on the appointment list level are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

co-authors on publications of the moving researchers. This is not the case, casting doubt on the alternative explanation that the effect is driven by researchers working in larger academic teams.

3.4.3 Returns to Tenure?

Do these results reflect genuine increases in publication activity or do they reflect other time-varying shocks? For example, the effects could be driven by the fact that as full professors, researchers in the natural sciences typically lead a research group and their

own "lab". To answer this question, in Table 3.4 we analyze the heterogeneity of the effect with respect to the type of position. Column (1) repeats the baseline specification, using citation-weighted publications as the dependent variable. In columns (2) and (3), we split the sample according to whether the position is a full chair (payment groups "C4" or "W3") or not (all other groups). Interestingly, the effect is not driven by researchers who move to a fully chaired position. This speaks against an explanation that moving researchers merely publish more because they get credit for heading their research group. In columns (4) and (5), we split the dependent variable into whether the researcher was the last author on the publication or not. In the natural sciences, being last author typically reflects credit for being the leader of a research group and the most senior author on a publication. We find that the effects are driven by an increase in publications where the researcher is *not* the last author on the publication. This again speaks against seniority as an explanation of the impact of moving on academic productivity. Finally, columns (6) and (7) split the number of quality-weighted publications into those on which the researcher is and those where she is not the first author. Being first author typically reflects having done the main lab work in the natural sciences. Because researchers in the sample apply to tenured positions, we would expect the effect of moving to be driven by publications where the researchers is not the first author. This is what we find.

In summary, this table shows that the effect is not merely driven by researchers moving to positions in which they hold the capital that the research is based on. It shows that the effect is based on tenured but junior positions and researchers who are not the most senior, but also not the most active researchers on their publications. Overall, it rather suggests an interaction effect of moving with age (Jones, 2010; Jones and Weinberg, 2011) than returns to seniority, which is in line with localized knowledge spillovers as the primary mechanism. These effects therefore provide some credibility to the effects found earlier.

Figure 3.4: Heterogeneity of Effect Across Fields

(a) Natural Sciences

(b) Other Fields
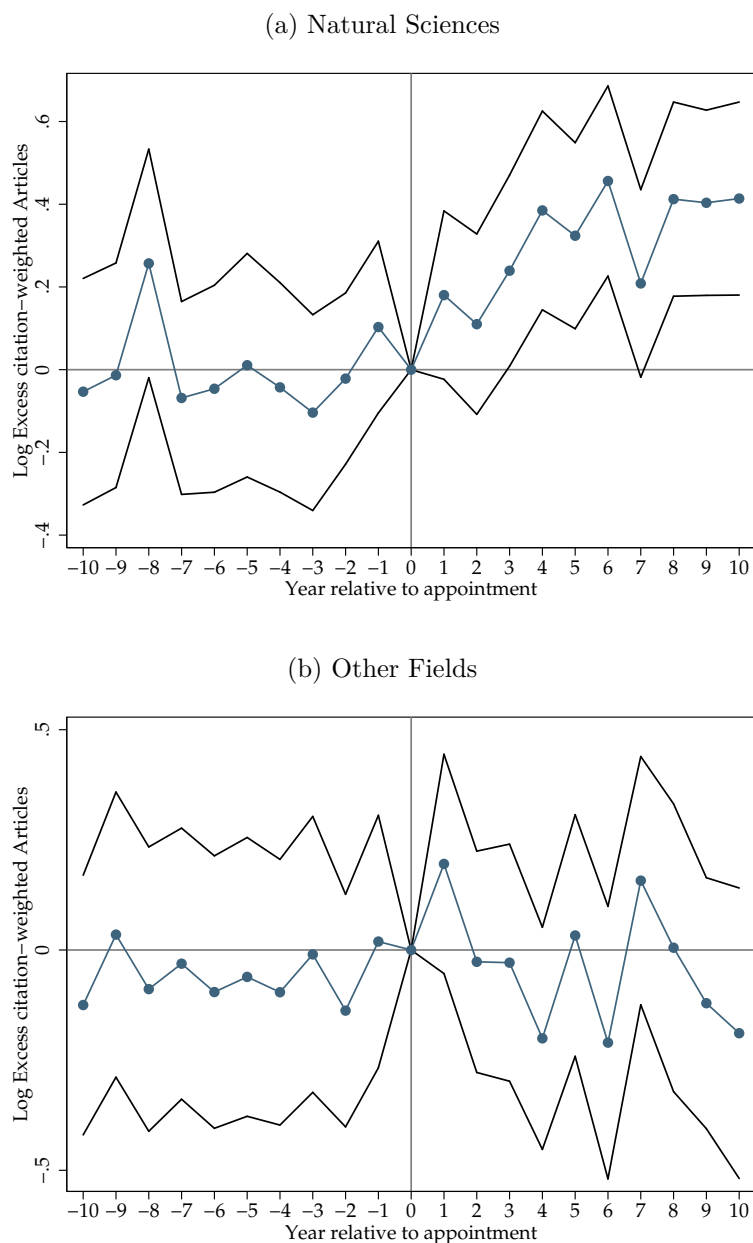
Notes: These graphs show the estimated number of yearly log excess publications by moving researchers, relative to the year of the move. Panel (a) limits the sample to researchers in the natural sciences, namely researchers in Biology, Chemistry and Pharmacy, Medicine, Physics, and Veterinary Medicine. Panel (b) limits the sample to researchers in all other fields, such as Business Administration, Philosophy, Theology, or Social Sciences. Publications are weighted by the number of authors and by the number of follow-on citations. To arrive at these estimates, we regress the yearly log number (plus one) of author- and citation- weighted publications by researchers on the appointment lists on an indicator variable equal to one if the researcher moved (interacted with time dummies), year fixed effects, and appointment list fixed effects. The dark blue line represents the 90% confidence bands for the estimated coefficients. Standard errors allow for clustering at the appointment list level. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

Table 3.4: Excess Publications of Moving Scientists: Heterogeneity of Effects along Position and Team Structure

Dep. Var.:	Ln(Citation-Weighted Publications+1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mover	0.07	0.19**	-0.02	-0.06	0.13*	-0.02	0.09*
	(0.07)	(0.09)	(0.10)	(0.04)	(0.07)	(0.06)	(0.05)
Post	-0.03	0.09**	-0.12**	0.00	-0.03	-0.26***	0.23***
	(0.04)	(0.04)	(0.06)	(0.01)	(0.04)	(0.04)	(0.04)
Post x Move	0.19***	0.15*	0.22***	0.03	0.16***	-0.05	0.23***
	(0.06)	(0.09)	(0.08)	(0.02)	(0.06)	(0.06)	(0.06)
Constant	1.87***	1.77***	1.96***	0.13***	1.74***	1.09***	0.79***
	(0.03)	(0.03)	(0.04)	(0.01)	(0.03)	(0.03)	(0.02)
List FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Baseline	Chair	No Chair	Last author	Not last author	First author	Not first author
Adj R2	0.45	0.52	0.38	0.18	0.44	0.20	0.35
Lists	317	142	175	317	317	317	317
Observations	10000	4470	5530	10000	10000	10000	10000

Notes: This table shows fixed effects regressions on the author-by-year level. The data comprises the five years before and five years after the appointment. The dependent variable the natural logarithm (plus one) of author- and citation-weighted publications of authors on the appointment list. In columns (2) and (3), we split the sample into researchers applying for chaired positions (pay scale C4 or W3) and researchers applying for other positions. In columns (4) and (5), we split the dependent variable into publications where the author is and is not the last-named author. In columns (6) and (7), we split the dependent variable into publications where the author is and is not the first-named author. Move is an indicator if the researcher was appointed, Post is an indicator for the post-move period and the indicator of interest PostxMove is the interaction for the period after the move of treated authors. In every regression we control for fixed effects for the appointment list under consideration. The constant reflects the intercept that makes the prediction calculated at the means of the independent variables equal to the mean of the dependent variable. Standard errors allowing for clustering on the appointment list level are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

3.4.4 Robustness Checks

The impact of moving on academic publications is identified if in the absence of the move, the publication rates of movers and non-movers would have developed in parallel (*parallel trends assumption*). While this assumption is inherently untestable, in this subsection we provide evidence that it is plausible in the context of this study.

Figure 3.5 shows estimates for time-varying coefficients where instead of using the actual treatment, we assign a treatment indicator to the highest ranked non-mover who is then compared to all other non-movers on the same appointment list. The contrast to the figures showing the actual treatment effects is striking: Besides one single uptick at the end of the sample period, there is no difference in the yearly publication rates between higher- and lower-ranked non-movers. This also speaks against the effect being driven by a mere time trend that is different between higher- and lower-ranked researchers, e.g., due to being responsible for a research group at a certain time of a researcher's career. If this was the case, we should see increases in the academic productivity of higher- vs. lower-ranked non-movers. This is not the case.

To quantify these estimates, in Table 3.5 we conduct placebo analyses in the difference-in-differences framework specified above. In the first column, we repeat the baseline estimates, again using citation-weighted publications as the dependent variable. Column (2) then shows the placebo treatment. In line with the identification assumption, the impact of "moving" is not significantly different from zero. In column (3), we use highest rank instead of actual move as treatment indicator, essentially estimating an intent-to-treat effect. While the coefficient estimate is positive, movers are not significantly more productive than non-movers.[14] Column (4) uses the subsample of lists for which the highest-ranked candidate rejected the offer.[15] The point estimate is even larger than the baseline estimate and is significantly different from zero at the five percent level. This again casts doubt on the alternative interpretation that first-ranked authors are

---

[14]To investigate this further, in Figure C.1 in the Appendix we show that while the time-varying treatment effects are estimated with more error, the pattern of coefficients is similar to the baseline pattern. However, the impact only arises around four years after the move, leading to a small coefficient estimate in the difference-in-differences specification in the table which only relies on the five years before and after the move.

[15]Therefore, the control group in this subsample includes the highest ranked candidate on each list.

Figure 3.5: Excess Publications of Non-Movers
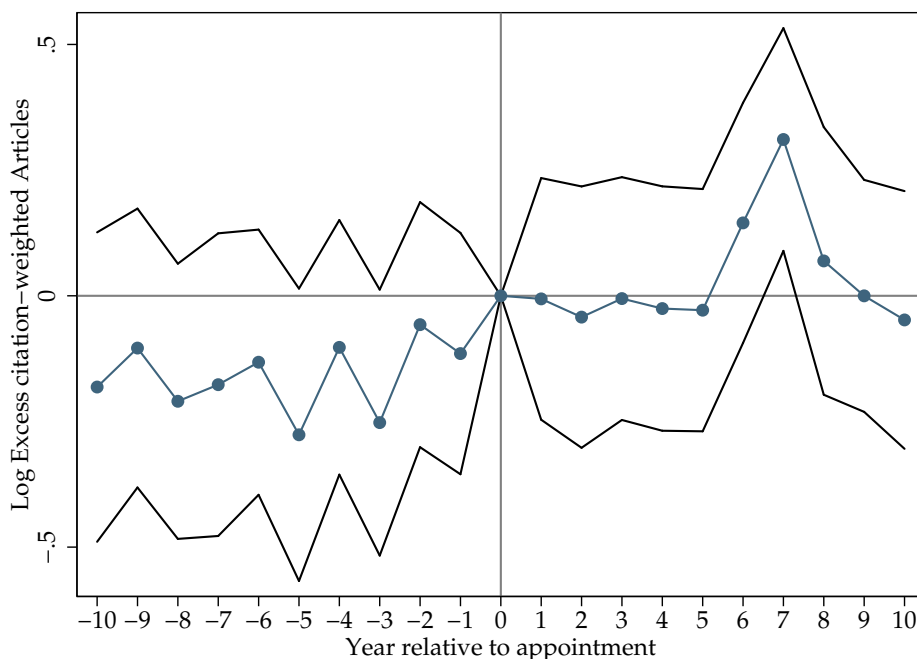
Notes: This graph shows the estimated number of yearly log excess publications by highest-ranked non-moving researchers, relative to the year of the move. Publications are weighted by the number of authors and by the number of follow-on citations. To arrive at these estimates, we regress the yearly log number of weighted publications by researchers on the appointment lists (plus one) on an indicator variable equal to one if the researcher is the highest-ranked non-moving researcher on the appointment list (interacted with time dummies), year fixed effects, and appointment list fixed effects. The dark blue line represents the 90% confidence bands for the estimated coefficients. Standard errors allow for clustering at the appointment list level. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

just inherently more productive in expectations than the other researchers on the same appointment list. It also casts doubt on the explanation that the effect merely stems from increased access to research funds or higher wages: Usually, declining an offer results from renegotiation with the origin university. As the highest ranked researcher is part of the counterfactual in this subsample, this would work against finding an effect. Column (5) shows that while researchers in medicine are an important and the largest single part of the sample, the effect is still significantly different from zero when estimated using only lists in fields other than medicine.

An important question is whether this effect is university-specific or whether it is generally present. While we cannot exclude that the effect is specific to this university, column

(6) uses publicly available data from the German magazine "Forschung und Lehre" (FuL) which regularly provides information on appointment offers and whether they were accepted or declined. Leveraging on rejected offers by analyzing repeated reports of the same position, we can estimate the impact of mobility on productivity for a subset of positions without restricting the sample to one university. The results show that the effect is similar to our baseline estimates. However, (second-ranked) movers are significantly less productive than the (first-ranked) non-movers in this sample, justifying our stringent data restrictions and stressing the importance of the detailed information about the appointment lists that make our setup unique.

Table 3.5: Robustness Analyses

Dep. Var.:	Ln(Citation-Weighted Publications+1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Mover	0.07	0.08	0.17**	-0.17	0.03	-0.38***
	(0.07)	(0.11)	(0.07)	(0.12)	(0.09)	(0.11)
Post	-0.03	-0.03	0.01	-0.02	0.05	-0.38***
	(0.04)	(0.06)	(0.04)	(0.08)	(0.04)	(0.07)
Post x Move	0.19***	0.01	0.09	0.30**	0.11*	0.16**
	(0.06)	(0.08)	(0.06)	(0.12)	(0.07)	(0.07)
Constant	1.87***	1.80***	1.84***	2.00***	1.75***	2.82***
	(0.03)	(0.07)	(0.03)	(0.06)	(0.03)	(0.06)
List FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Baseline	Non-Movers	ITT	Rejecters	W/o Medicine	FuL
Adj R2	0.45	0.48	0.45	0.51	0.50	0.41
Lists	317	317	317	56	196	147
Observations	10000	6440	10000	2070	6180	3400

Notes: This table shows fixed effects regressions on the author-by-year level. The data comprises the 5 years before and 5 years after the appointment. The dependent variable the natural logarithm (plus one) of author- and citation-weighted publications of authors on the appointment list. In column (2), we assign the move-indicator to the highest ranked non-mover on each appointment list. In column (3), we assign the move-indicator to the highest ranked researcher on the sample, irrespective of the actual move, thus estimating an intent-to-treat effect. In column (4), we use the subset of lists where the highest ranked researcher did not accept the offer. In column (5), we use the subset of lists which are not in the field of medicine. In column (6), we use rejected offers which were publicly announced in the magazine "Forschung und Lehre" (FuL). Move is an indicator if the researcher was appointed (in all columns but 2), Post is an indicator for the post-move period and the indicator of interest PostxMove is the interaction for the period after the move of treated authors. In every regression we control for fixed effects for the appointment list under consideration. Standard errors allowing for clustering on the appointment list level are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005 in columns (1) through (5) and from FuL in column (6). Publication data stems from Scopus.

In summary, this table shows that the identification assumption of parallel trends in

quality-weighted publications between movers and non-movers in the absence of the move is plausible. It is also indicative that the alternative explanation of increased research funds is unlikely to be driving the results and that the effect is not restricted to this university. Overall, moving to a new university seems to increase the number of yearly citation-weighted publications.

3.5   Conclusion

This chapter provides novel results on the question whether labor mobility of scientists increases their academic productivity. We use data on newly hired university scientists and their runners-up for the same academic position at a large university in Germany (Watzinger et al., 2017b). The setup provides important advantages over previous studies of the same question: All scientists on the appointment lists applied for the same position, were extensively vetted, and were found acceptable by the hiring committee. The setup therefore circumvents the inherent problems of finding an appropriate control group in the presence of self-selection of researchers into moving to other universities (Roy, 1951; Azoulay et al., 2017; Ganguli, 2015a).

In response to the move, the scientific productivity of movers relative to non-movers on the same appointment list increases by around 13% as measured by (citation-weighted) publications. This translates to around 0.5 more author- and citation-weighted publications per year. The effect is driven by the natural sciences. In contrast to previous findings, the effect stems from researchers on lists with above-median pre-move academic impact. It is smaller for researchers applying for full chairs and is not driven by publications on which the moving researcher is the last author. In line with Catalini (forthcoming) and Boudreau et al. (forthcoming), the observed increase in productivity points towards the existence of search costs in knowledge space. The effect is only present for articles in the top and the bottom third of the citation distribution, not for "average" articles. Therefore, experimentation costs seem to have decreased as well, similar to the findings in Catalini (forthcoming) on the impacts of relocating offices at a Paris university.

In a variety of robustness checks, we show that the identification assumption of parallel

trends in scientific productivity is plausible. For example, leveraging our event-study approach, we show that there is no differential pre-trend for movers. What is more, among non-movers, higher rank on the appointment list is unrelated to higher (post-move) productivity. Finally, the treatment effect is similar (and even slightly larger) for the subsample where the highest-ranked researcher declined the offer. This result also speaks against an interpretation where the impact is only generated by increased access to research funds. The reason is that scientists who decline offers mostly do so after renegotiation with their home university, which should work against finding any effect. However, a more thorough investigation into alternative potential mechanisms is still necessary. The results in this chapter should therefore be regarded as preliminary.

Because researchers move for a reason, the ATT we estimate is likely larger than the treatment effect for scientists who are not interested in moving to the university. A more thorough investigation into the transferability of these results is necessary, but left for future research. Thus, one should keep in mind that the size of our estimates does not necessarily translate to an average treatment effect of moving among the general population of scientists.

Yet, the results of this chapter show that scientist mobility may indeed have benefits over and beyond generating spillovers for other researchers (Watzinger et al., 2017b). This largely neglected impact of mobility on academic output seems important: The size of the effects suggests that labor mobility may be a fruitful way to increase scientific productivity for researchers. Our estimates are in line with recent findings on the benefits of (international) scientist mobility for academic productivity (Franzoni et al., 2014; Ganguli, 2015a). Finally, these estimates also explain why researchers are willing to face the substantial costs of moving, even in the absence of market mechanisms which internalize the spillover effects of scientist mobility.

Chapter 4

Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness

4.1 Introduction

How do alternative job opportunities affect teacher quality? This is a crucial policy question as teachers are a key input in the education production function (Hanushek and Rivkin, 2012) who affect their students' outcomes even in adulthood (Chetty et al., 2014b). Despite their importance, individuals entering the teaching profession in the United States tend to come from the lower part of the cognitive ability distribution of college graduates (Hanushek and Pace, 1995). One frequently cited reason for not being able to recruit higher-skilled individuals as teachers is low salaries compared to other professions (e.g., Dolton and Marcenaro-Gutierrez, 2011; Hanushek et al., 2014).

Existing research provides evidence consistent with the argument that outside options matter. A first strand of the literature has used regional variation in relative teacher salaries, finding that pay is positively related to teachers' academic quality (e.g., Figlio, 1997). A second strand has used long-run changes in the labor market – in particular, the expansion of job opportunities for women – finding that the academic quality of new

---

This chapter is based on joint work with Marc Piopiunik and Martin West (Nagler et al., 2015).

teachers is lower when job market alternatives are better (e.g., Bacolod, 2007). However, both bodies of evidence suffer from key limitations. First, relative pay may be endogenous to teacher quality. Second, measures of academic quality are poor predictors of teacher effectiveness (cf. Jackson et al., 2014). This important policy question therefore remains unresolved.

We exploit business cycle conditions at career start as a source of exogenous variation in the outside labor-market options of potential teachers.[1] Because the business cycle conditions at career start are exogenous to teacher quality, our reduced-form estimates reflect causal effects. In contrast to prior research, we directly measure teacher quality with value-added measures (VAMs) of impacts on student test scores, a well-validated measure of teacher effectiveness (e.g., Kane and Staiger 2008; Chetty et al. 2014a,b; see Jackson et al. 2014 for a review). Combining our novel identification strategy with VAMs for individual elementary school teachers from a large U.S. state, we provide causal evidence on the importance of alternative job opportunities for teacher quality.

Our value-added measures are based on individual-level administrative data from the Florida Department of Education on 33'000 4th- and 5th-grade teachers in Florida's public schools and their students. The data include Florida Comprehensive Assessment Test (FCAT) math and reading scores for every 3rd-, 4th-, and 5th-grade student tested in Florida in the 2000-01 through 2008-09 school years. The data also contain information on teachers' total experience in teaching (including experience in other states and private schools), which is used to compute the year of entry into the profession (which is not directly observed). Following Jackson and Bruegmann (2009), we regress students' math and reading test scores separately on their prior-year test scores, student, classroom, and school characteristics, and grade-by-year fixed effects to estimate each teacher's value-added. We then relate the VAMs in math and reading to several business cycle indicators from the National Bureau of Economic Research (NBER) and the Bureau of Labor Statistics (BLS).

We find that teachers who entered the profession during recessions are roughly 0.10 stan-

---

[1]To our knowledge, the idea that outside labor-market options at career start matter for teacher quality was first proposed by Murnane and Phillips (1981) in their classic paper on "vintage effects." Zabalza (1979) provides early evidence that starting salaries within teaching influence individual decisions to enter the profession, while Dolton (1990) finds large impacts of teachers' relative earnings and earnings growth.

dard deviations (SD) more effective in raising math test scores than teachers who entered the profession during non-recessionary periods. The effect is half as large for reading value-added. Quantile regressions indicate that the difference in math value-added between recession and non-recession entrants is most pronounced at the upper end of the effectiveness distribution. Based on figures from Chetty et al. (2014b), the difference in average math effectiveness between recession and non-recession entrants implies a difference in students' discounted life-time earnings of around $13'000 per classroom taught each year.[2] Under the more realistic assumption that only 10% of recession-cohort teachers are pushed into teaching because of the recession, these recession-only teachers are roughly one SD more effective in teaching math than the teachers they push out. Based on the variation in teacher VAMs in our data, being assigned to such a teacher would increase a student's test scores by around 0.20 SD.

Placebo regressions show that neither business cycle conditions in the years before or after teachers' career starts, nor those at certain critical ages (e.g., when most students enter or complete college), impact teacher effectiveness; only conditions at career start matter. Nor are our results driven by differential attrition among recession and non-recession cohorts. Although teachers entering during recessions are more likely to exit the profession, the observed attrition pattern works against our finding and suggests that our results understate the differences in effectiveness between recession and non-recession cohorts at career start. The results are also not driven by any single recession cohort, but appear for most recessions covered by our sample period. Using alternative business cycle measures such as unemployment levels and changes yields very similar results. The recession effect is not driven by differences in teacher race, gender, age at career start, cohort sizes, or school characteristics. Our finding that the effect of recessions on teacher effectiveness is twice as strong in math as in reading is consistent with evidence that wage returns to numeracy skills are twice as large as those to literacy skills in the U.S. labor market (Hanushek et al., 2015). These results are also consistent with the common finding that students' reading scores are more difficult to improve than their math scores

---

[2]Chetty et al. (2014b) estimate that students who are taught by a teacher with a 1 SD higher value-added measure at age 12 earn on average 1.3% more at age 28. Assuming a permanent change in earnings and discounting life-time earnings at 5%, this translates into increases in discounted life-time earnings of $7'000 per student. We obtain our estimate by multiplying this number by our effect size and average classroom size.

(cf. Jackson et al., 2014).

To motivate our analysis, we present a stylized Roy model (Roy, 1951) in which more high-skilled individuals choose teaching over other professions during recessions because of lower (expected) earnings in those alternative occupations. The model's main assumption is that teaching is a relatively stable occupation over the business cycle. This seems reasonable since teacher demand depends primarily on student enrollment and is typically unresponsive to short-run changes in macroeconomic conditions (e.g., Berman and Pfleeger, 1997). We present evidence that supports our interpretation of these results as supply effects, rather than demand effects or direct impacts of recessions on teacher effectiveness.[3]

Consistent with this model, existing studies show that the supply of workers for public sector jobs in the U.S. is higher during economic downturns (e.g., Krueger, 1988; Borjas, 2002). Falch et al. (2009) document the same pattern for the teaching profession in Norway. Teach For America, an organization that recruits academically talented college graduates into teaching, saw a marked decline in the number of qualified applicants during the recent economic recovery (New York Times, 2015). Meanwhile, several U.S. states have reported sharp declines in enrollment in university-based teacher preparation programs as the job market has improved (National Public Radio, 2015).

Our results have important policy implications. First, they suggest that increasing the economic benefits of becoming a teacher may be an effective strategy to increase the quality of the teaching workforce. In contrast to de Ree et al. (forthcoming), who find that unconditional increases in teacher pay for incumbent teachers do not improve student achievement, our results suggest that *selection into teaching* is affected by changes in economic benefits. This is in line with field-experimental evidence from developing countries: For example, Ashraf et al. (2016) find that selecting individuals who care about career incentives rather than those who are intrinsically motivated leads to better outcomes in

---

[3]Figure 4.1 confirms that employment in the private sector is much more cyclical than employment in (state and local) education. The major exception is the recession period of 1980-1982, but our results for this recession differ from and work against our main findings. Kopelman and Rosen (2016) report higher job security for public sector jobs (including teaching) than for jobs in the private sector. Consistently, newspapers have reported that teaching is recession-proof. During the most recent recession, job security for teachers did decline substantially (e.g., New York Times, 2010). This last downturn does not drive our results.

public service delivery. Second, our results also suggest that recessions may provide a window of opportunity for the public sector to hire more able applicants. Finally, they also suggest that recent improvements in cognitive skills among new teachers in the U.S. documented by Goldhaber and Walch (2013) may be attributable to the 2008-09 financial crisis, rather than an authentic reversal of long-term trends.

We extend previous research that has called attention to the potential importance of outside job options for teacher quality. Most recently, Britton and Propper (2016) exploit centralized wage regulation that generates regional variation in teachers' relative wages in England to document positive effects of relative teacher pay on school productivity.[4] However, their school-level data do not allow them to disentangle selection into the teaching profession from the sorting of teachers into specific schools and potential differences in teacher effort due to efficiency wage effects. Bacolod (2007) documents a decrease in the academic quality (as measured by standardized test scores and undergraduate institution selectivity) of female teachers in the U.S. over time that coincided with improvements in women's outside options.[5] In comparison with her study, we use a more rigorous identification strategy and direct measures of teachers' performance on the job. Our study is therefore the first to document a causal effect of outside labor-market options on the effectiveness of entering teachers in raising student test scores.

Business cycle fluctuations have previously been exploited as a strategy to identify selection effects in the labor market. Oyer (2008), for example, studies the impact of the business cycle on the likelihood that MBA graduates enter the banking sector.[6] Boehm and Watzinger (2015) show that PhD economists graduating during recessions are more productive in academia, a finding best explained by a Roy-style model. While these studies enhance the plausibility of our findings, they relate to rather small groups in the labor market with highly specialized skills. Teachers, in contrast, make up roughly 3 percent of full-time workers in the U.S. and play a critical role in developing the human capital of future generations. Moreover, little is known about how to improve the quality of the

---

[4]Loeb and Page (2000) similarly relate regional variation in relative teacher wages and unemployment rates to rates of educational attainment but also lack direct measures of teacher quality.

[5]Corcoran et al. (2004), Hoxby and Leigh (2004), and Lakdawalla (2006) provide additional evidence of the importance of outside job options for the supply of American teachers.

[6]A small literature also documents persistent negative wage effects of completing college during a recession (e.g., Kahn, 2010; Oreopoulos et al., 2012).

teaching workforce. Thus, extending this identification strategy to teacher quality fills an important gap in the literature.

The chapter proceeds as follows. Section 4.2 presents a simple model of occupational choice. Section 4.3 briefly describes the teaching profession in Florida, introduces the data, explains our value-added measures, and presents our empirical model. Section 4.4 reports results on the relationship between business cycle conditions at career start and teacher effectiveness in math and reading and provides robustness checks. Section 4.5 discusses potential implications for policymakers. Section 4.6 concludes.

4.2 A Simple Model of Occupational Choice

To motivate our analysis, we present a simple Roy-style model of self-selection (Roy, 1951) where individuals choose an occupation to maximize (expected) earnings.[7] Specifically, individuals can choose between working in the teaching sector (t) and working in the business sector (b), which represents all outside labor-market options of potential teachers. Earnings depend on average earnings in the respective sector, μ, and the individual's ability, v. Hence, earnings in the two sectors for any individual with ability v can be written as follows:

$$w_t = \mu_t + \eta_t v$$

$$w_b = \mu_b + v - s$$

where w_t and w_b are earnings in the teaching and business sector, respectively; v is the (uni-dimensional) ability of the individual, distributed with mean zero and standard deviation σ_v^2; and η_t denotes the relative returns to ability in teaching versus business. If ability is valued both in business and teaching, but teaching has lower returns to ability, then $\eta_t \in (0,1)$.[8] If there are no returns to ability in teaching, then $\eta_t = 0$.[9]

---

[7]Individuals may, of course, be motivated by other concerns than earnings. One can therefore think of earnings as a proxy for lifetime utility.

[8]Wages are more compressed in the government-dominated teaching profession than in the business sector (cf. Hoxby and Leigh, 2004; Dolton, 2006).

[9]Since our model only uses one dimension of ability, we implicitly assume that the two abilities typically used in Roy models are positively correlated (i.e., $\eta_t \geq 0$). We make this assumption for expositional

The term s (≥ 0) represents the reduction in (expected) earnings in the business sector *relative* to the reduction in earnings in the teaching sector (which is normalized to zero) during recessions. The model thus allows for recessions to affect earnings in the teaching profession, but assumes that the impact is stronger in the business sector. Empirically, employment in the teaching sector is less cyclical than employment in the business sector (see Figure 4.1; see also Berman and Pfleeger 1997; Simpkins et al. 2012).

Figure 4.1: Employment in Private Sector and Local and State Education
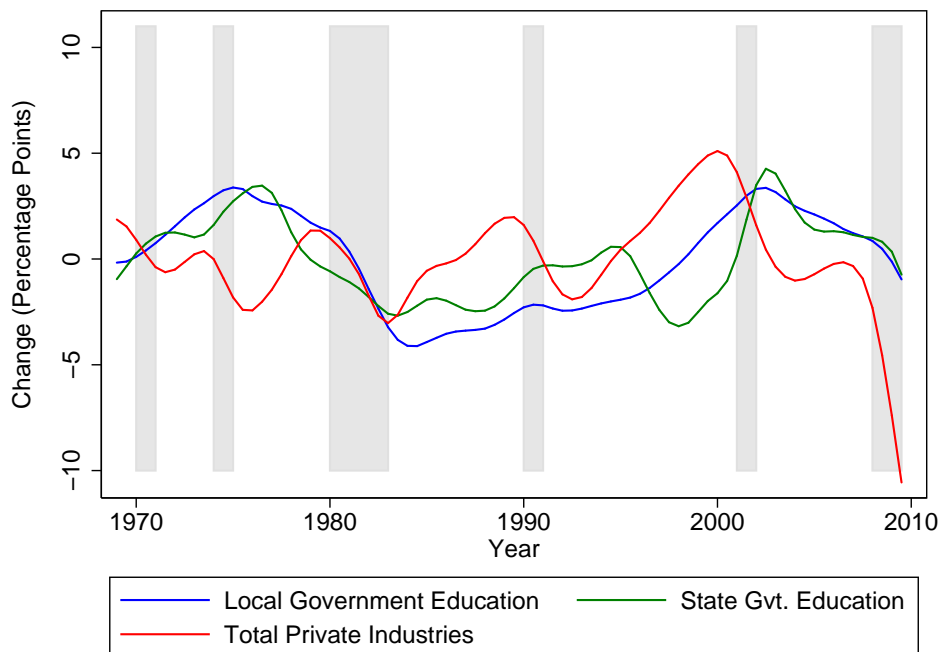
Notes: Data come from the Current Employment Statistics (Establishment Survey) of the U.S. Bureau of Labor Statistics as compiled by the Federal Reserve Bank of St. Louis. Number of employees in the indicated sector are seasonally adjusted. Semiannual frequency, indexed to 100 in second half of 2007, and detrended. Shaded areas: Recessions as defined by the NBER.

Individuals choose teaching if $w_t > w_b$, which is equivalent to $v < \frac{\mu_t - \mu_b + s}{1 - \eta_t}$. Hence, the share of individuals seeking employment in the teaching sector is given by

$$Pr(t) = Pr\left(v < \frac{\mu_t - \mu_b + s}{1 - \eta_t}\right) = F\left(\frac{\mu_t - \mu_b + s}{1 - \eta_t}\right)$$

where $F(\cdot)$ is the cumulative distribution function of individuals' ability v, which is con-

---

clarity only, but note that it has empirical support. For example, Chingos and West (2012) show that, among 35,000 teachers leaving Florida public schools for other industries, a 1 SD increase in teacher value-added is associated with 6–8 percent higher earnings in non-teaching jobs.

tinuously distributed over $\mathbb{R}$. If $0 \leq \eta_t < 1$, recessions increase the supply and (average) quality of potential teachers. When a recession hits the economy (increasing s), the share of individuals seeking employment in the teaching sector increases because the earnings of teachers increase relative to more cyclical outside options:

$$\frac{\partial Pr(t)}{\partial s} = f\left(\frac{\mu_t - \mu_b + s}{(1 - \eta_t)}\right)\frac{1}{1 - \eta_t} > 0.$$

The average ability of individuals seeking employment in teaching increases because individuals with higher ability prefer working in the teacher profession; formally, $\frac{\partial v_{marg}}{\partial s} = \frac{1}{(1-\eta_t)} > 0.$[10] We expect our empirical analysis to be consistent with this prediction as the underlying assumptions (i.e., $\eta_t \in (0,1)$ and $s \geq 0$) have strong empirical support. If $\eta_t > 1$, we would expect to find negative effects of recessions on teacher quality.

Empirically, we analyze the importance of outside labor-market options for teacher quality. In our model, changes in labor-market opportunities are modeled as changes in expected earnings. Both employment probability and relative earnings likely change in favor of the teaching profession during recessions, but we cannot discriminate between these two channels in our empirical analysis. If the model's assumptions hold, however, our estimates shed light on whether increasing teacher pay would increase teacher quality.

While our simple model only addresses the supply of teachers, fluctuations in demand could in theory also explain changes in teacher quality over the business cycle. Fluctuations in demand would lead to higher quality of teachers entering during recessions if the following two conditions hold. First, school authorities are able to assess the quality of inexperienced applicants and accordingly hire the more able ones. Second, the number of hired teachers is smaller during recessions than during booms. If either of these two conditions does not hold, fluctuations in demand would not cause recession teachers to be more effective than non-recession teachers. We return to this issue after presenting our main results.

---

[10]Marginal individuals, indifferent between working in the teaching sector and working in the business sector, are characterized by $v_{marg} = \frac{\mu_t - \mu_b + s}{(1-\eta_t)}$.

4.3   Setting, Data, and Empirical Strategy

First, we document the feasibility of a short-run response in teacher supply to fluctuations in economic conditions by providing information on the pool of potential teachers nationally and describing the requirements for entry into the teaching profession in Florida. Second, we introduce the data and describe our empirical strategy. We use variation in career start years to analyze the impact of outside labor-market opportunities on the selection into teaching. We estimate the career start year by subtracting total experience in teaching from the year in which we observe the teacher. Third, we describe our empirical strategy, including the construction of our value-added measures of teacher effectiveness.

4.3.1   Supply of Potential Teachers in Florida

Nationally, the number of individuals completing teacher education programs each year has been roughly double the number of newly hired teachers since at least 1987, when the earliest comprehensive data are available (Cowan et al., 2016). This implies that, at any point in time, there is a large pool of potential teachers nationally who are eligible to obtain certification immediately, regardless of the rigidity of state certification regimes. It also suggests that, for many potential teachers, the key decision about whether or not to enter the profession occurs when they enter the labor market rather than when they choose a degree program.

Contrary to the national data, the demand for new teachers in Florida has exceeded the supply of new graduates from in-state preparation programs since at least the 1980s due to growth in the student population and, since 2003, a statewide class-size reduction mandate (Moe, 2006). In response to this pressure, state policymakers have consistently sought to recruit teachers from outside Florida. For example, a 1983 law required the Florida Department of Education to create a teacher referral and recruitment center to pursue strategies such as advertising teaching positions in states with declining enrollments and in major newspapers and establishing a national toll-free number to handle inquiries from prospective teachers (Florida Department of Education, 1986). In the 1980s, the state estimated that as many as 45 percent of new teachers in Florida had completed

their preparation program in another state. Similarly, the U.S. Department of Education (2013) indicates that 23 percent of individuals receiving their initial Florida teaching credential in 2009 were prepared out-of-state. In our data, 19% of teachers report having teaching experience in other states, providing a lower bound on the number who prepared elsewhere. These statistics highlight the extent to which the pool of potential teachers for Florida public schools is national in scope and therefore apt to be influenced by national rather than state-specific economic conditions.

Temporary fluctuations in economic conditions are also more likely to influence selection into teaching when certification regimes permit as many individuals as possible to enter the profession without completing additional training. Traditionally, American states required potential teachers to complete an undergraduate or master's degree teacher preparation program in order to be certified to teach. Although in practice individuals without certification were often granted emergency credentials, these certification requirements likely constrained any short-term supply response. In recent decades, however, shortages of certified teachers in specific subject areas led many states to create alternative entry routes that allow college graduates who have not completed a traditional preparation program to begin teaching immediately while completing the remaining requirements for professional certification. As of 2011, 45 states had approved an alternative certification program and individuals completing these programs comprised roughly 20 percent of all individuals completing teacher preparation programs nationwide (U.S. Department of Education, 2013).

Florida's certification regime is typical of those states that have created alternative entry routes into teaching. The state initially awards professional teaching certificates only to graduates of state-approved teacher preparation programs who have passed tests of general knowledge, professional education, and the subject area in which they will teach.[11] However, college graduates who have not completed a teacher preparation program are eligible for a temporary certificate if they majored or completed a specified set of courses in the relevant subject area. They may also become eligible for a temporary certificate by passing a test of subject-matter knowledge. Individuals with a temporary certificate may

---

[11]Florida also recognizes professional certificates in comparable subject areas granted by other states and by the National Board of Professional Teaching Standards.

then teach for up to three years while completing 15 credit hours of education courses and a school-based competency demonstration program. These arrangements allow any college graduate to enter the teaching profession in Florida (at least temporarily) in response to labor market conditions by passing a single exam.

Florida first authorized alternative certification for teachers in all grades and subject areas in 1997 and, since the 2002-03 school year, has required that each school district in the state offer its own alternative certification program (Moe, 2006). However, the state permitted school districts to hire teachers on temporary certificates for up to two years even before creating a formal alternative route and, until 1988, allowed the same individual to receive a temporary certificate multiple times (Florida Department of Education, 1986). The extent to which certification requirements may have constrained the supply response to labor market conditions among college graduates in the state prior to that period is therefore unclear.

4.3.2 Administrative Data from the State of Florida

Teacher value-added measures are based on administrative data from the Florida Department of Education's K–20 Education Data Warehouse (EDW). Our EDW data include observations of every student in Florida who took the state test in the 2000–01 through 2008–09 school years, with each student linked to his or her courses (and corresponding teachers). We focus on scores on the Florida Comprehensive Assessment Test (FCAT), the state accountability system's "high-stakes" exam. Beginning in 2001, (only) students in grades 3–10 were tested each year in math and reading. Thus annual gain scores can be calculated for virtually all students in grades 4–10 starting in 2002. The data include information on the demographic and educational characteristics of each student, including gender, race, free or reduced price lunch eligibility, limited English proficiency status, and special education status.

The EDW data also contain detailed information on individual teachers, including their demographic characteristics and teaching experience. We use only 4th- and 5th-grade teachers because these teachers typically teach all subjects, thus avoiding spillover effects from other teachers. We construct a dataset that connects teachers and their students

in each school year through course enrollment data. Our teacher experience variable reflects the total number of years the teacher has spent in the profession, including both public and private schools in Florida and other states. Because the experience variable contains a few inconsistencies, we assume the latest observed experience value is correct, and adjust all other values accordingly. Year of career start is defined as the calendar year at the end of the school year a teacher is observed in the data minus total years of teaching experience.[12] Starting from the baseline dataset that contains all 4th- and 5th-grade students with current and lagged test scores, we apply several restrictions to keep only those teachers who can be confidently associated with students' annual test score gains. We only keep student-teacher pairs if the teacher accounts for at least 80% of the student's total instruction time (deleting 24.5% of students from the baseline dataset). We exclude classrooms that have fewer than seven students with current and lagged scores in the relevant subject and classrooms with more than 50 students (deleting 1.8% of students). We also drop classrooms where more than 50% of students receive special education (deleting 1.5% of students). We further exclude classrooms where more than 10% of students are coded as attending a different school than the majority of students in the classroom (deleting 0.7%). Finally, we drop classrooms for which the teacher's experience is missing (deleting 1.8% of students). Our final dataset contains roughly 33'000 public school teachers with VAMs for math and reading.

Our main indicator for the U.S. business cycle is a dummy variable reflecting recessions as defined by the National Bureau of Economic Research (NBER). Recession start and end dates are determined by NBER's Business Cycle Dating Committee based on real GDP, employment, and real income. The NBER does not use a stringent, quantitative definition of a recession, but rather a qualitative one, defining a recession as "a period between a peak and a trough" (see http://www.nber.org/cycles/recessions.html). For example, the NBER dates the economic downturn of the early 1990s to have occurred between July 1990 (peak) and March 1991 (trough). We code our recession indicator variable to be one in 1990 (the beginning of the recession), and zero in 1991. Accordingly, teachers starting their careers in the 1990-91 school year are classified as having entered during a recession.

---

[12]We adjust career start dates for gaps in teaching observed after 2002, when we directly observe whether a teacher is working in Florida public schools each year. Results are very similar when using the original, uncorrected values.

In robustness checks, we use alternative business cycle indicators such as unemployment for college graduates (in levels and annual changes, nationwide and in Florida), overall unemployment for specific industries, and GDP, which come from the Bureau of Labor Statistics and the Bureau of Economic Analysis. NBER's recession indicator is highly correlated with unemployment rates (both levels and annual changes) and GDP.

4.3.3 Empirical Strategy

This section describes the estimation of teachers' value-added and our strategy for analyzing the relationship between business cycle conditions at career start and teacher value-added.

Estimating Teacher Value-Added

Teacher value-added measures (VAMs) aim to gauge the impact of teachers on their students' test scores. We estimate VAMs for 4th- and 5th-grade teachers based on students' test scores in math and reading from grades 3–5.[13] To estimate the value-added for each teacher, we regress students' math and reading test scores separately on their prior-year test scores, student, classroom, and school characteristics as well as grade-by-year fixed effects. Student-level controls include dummy variables for race, gender, free- and reduced-price lunch eligibility, limited English proficiency, and special-education status. Classroom controls include all student-level controls aggregated to the class level and class size. School-level controls include enrollment, urbanicity, and the school-specific shares of students who are black, white, Hispanic, and free- and reduced-price lunch eligible.

To obtain an estimate of each teacher's value-added, we add a dummy variable, θ_j, for each teacher:

$$A_{ijgst} = \widehat{\alpha} A_{i,t-1} + \beta X_{it} + \gamma C_{it} + \lambda S_{it} + \pi_{gt} + \theta_j + \epsilon_{ijgst}$$

---

[13]Note that student testing in Florida starts in grade 3 only.

where A_{ijgst} is the test score of student i with teacher j in grade g in school s in year t (standardized by grade and year to have a mean of zero and standard deviation of one); $A_{i,t-1}$ contains the student's prior-year test score in the same subject; X_{it}, C_{it}, and S_{it} are student-, classroom-, and school-level characteristics; π_{gt} are grade-by-year fixed effects; and ϵ_{ijgst} is a mean-zero error term. After estimating the teacher VAMs, θ_j, we standardize them separately for math and reading to have a mean of zero and a standard deviation of one.[14]

Since test scores suffer from measurement error, the coefficient on the lagged test score variable, $A_{i,t-1}$, is likely downward biased, which would bias the coefficients on other control variables correlated with lagged test scores. We therefore follow Jackson and Bruegmann (2009) and use $\widehat{\alpha}$, which is the coefficient on the lagged test scores from a two-stage-least-squares model where the second lag of test scores is used as an instrument for the lagged test scores (see the web appendix of Jackson and Bruegmann (2009) for details). Because this procedure requires two lags of test scores, the estimation of $\widehat{\alpha}$ is based on 5th-grade students only (students were not tested in grade 2).

Although widely used by researchers, the reliability of value-added models of teacher effectiveness based on observational data continues to be debated (see, e.g., Jackson et al., 2014; Rothstein, 2017). The key issue is whether non-random sorting of students and teachers both across and within schools biases the estimated teacher effectiveness. This would be the case if there were systematic differences in the unobserved characteristics of students assigned to different teachers that are not captured by the available control variables.[15]

Value-added models have survived a variety of validity tests, however. Most importantly, estimates of teacher effectiveness from observational data replicate VAMs obtained from experiments where students within the same school were randomly assigned to teachers

---

[14]To simplify notation, we drop the subscripts j, g, and s for the lagged test score and for the student-, classroom-, and school-level characteristics. We control for school characteristics rather than include school fixed effects because the latter would eliminate any true variation in teacher effectiveness across schools. However, we show below that our results are robust to the inclusion of both school and school-by-year fixed effects (Table D.2 in appendix D). We include grade-by-year fixed effects because test scores have been standardized using the full sample of students and because teachers are not observed in all years.

[15]For a more general discussion on the assumptions behind value-added models, see Todd and Wolpin (2003).

(Kane and Staiger, 2008; Kane et al., 2013). Chetty et al. (2014a) and Bacher-Hicks et al. (2014) exploit quasi-random variation from teachers switching schools to provide evidence that VAMs accurately capture differences in the causal impacts of teachers across schools. Using a different administrative data set, Rothstein (2017) argues that evidence on school switchers does not rule out the possibility of bias.

Even if our VAMs were biased by non-random sorting of students and teachers, however, it is unclear whether and, if so, in what direction this would bias our estimates of the relationship between recessions at career start and teacher effectiveness.

Finally, some critics argue that value-added measures may reflect teaching to the test rather than true improvements in knowledge. In a seminal study, Chetty et al. (2014b) find that having been assigned to higher value-added teachers increases later earnings and the likelihood of attending college and decreases the likelihood of teenage pregnancy for girls. Of course, there may be other dimensions of teacher quality not captured by VAMs (e.g., Jackson, 2012). The weight of the evidence, however, indicates that teacher value-added measures do reflect important aspects of teacher quality.

Business Cycle Conditions at Career Start and Teacher Value-Added

To estimate the effect of business cycle conditions at career start on teacher effectiveness, we relate the macroeconomic conditions in the U.S. during the career start year to a teacher's value-added in math and reading. Specifically, we estimate the following reduced-form model:

$$\widehat{\theta}_j = \alpha + \gamma Rec_{js} + \beta X_j + u_j$$

where $\widehat{\theta}_j$ is the value-added of teacher j (either in math or in reading). Rec_{js} is a binary indicator that equals 1 if teacher j started working in the teaching profession (in year s) in a recessionary period and equals 0 otherwise. The vector X_j includes teacher characteristics. Most importantly, it contains total experience in the teaching profession (yearly dummies up to 30 years of experience), which is not accounted for in the VAM computation but has been shown to influence teacher effectiveness (Papay and Kraft,

2015).[16] As experience differs between recession and non-recession teachers – due in part to the idiosyncratic distance between recessions and the time period covered by our administrative data – experience is a necessary control. Additional teacher characteristics included in some specifications are year of birth, age at career start, educational degree, gender, and race. Note that these teacher characteristics do not influence the business cycle. The reduced-form estimate γ (controlling only for experience) therefore identifies a causal effect. To the extent that the inclusion of additional controls changes the estimate of γ, they represent mechanisms rather than confounders. Because the source of variation is the yearly business cycle condition, we always adjust standard errors for clustering at the level of the career start year.

Based on our Roy model, we expect to find a positive effect of recessions at career start on teacher effectiveness since recessions negatively shock the outside options of potential teachers. Due to this shock, both the number and the average quality of applicants increases, leading to higher average value-added in recession cohorts. Since we do not observe the intermediate steps (e.g., application rates or earnings), we estimate a reduced-form relationship between teacher value-added and business cycle conditions at career start.

Critics of this model might argue that teacher effectiveness is unrelated to productivity in other occupations, but rather depends on intrinsic motivation. This should work against any positive effect of recessions on teacher value-added. At the margin, recession-only teachers should be less intrinsically motivated as they enter the teaching profession *because of low outside options*. Evidence of a positive effect would therefore also suggest that intrinsic motivation is of second-order importance relative to the effects of economic benefits through selection on ability (cf. Ashraf et al., 2016). Note also that because the effectiveness of all teachers in our sample is estimated during the same period (2001-2009), systematic differences in the effort levels of recession and non-recession teachers due to differences in the (policy or economic) environment seem unlikely.

---

[16]Previous work has shown that teacher experience affects teacher value-added non-linearly (e.g., Rockoff, 2004). Wiswall (2013) shows that non-parametric specifications yield the most convincing results. Our results are robust to using teachers with above 20 or 25 years of experience as the omitted category.

4.4 Business Cycle Conditions at Career Start and Teacher Effectiveness

We start by documenting differences in math and reading effectiveness between recession and non-recession teachers. Using kernel density plots and quantile regressions, we show at which parts of the effectiveness distribution recession and non-recession teachers differ. In placebo regressions, we show that teacher effectiveness is not associated with business cycle conditions several years before and after career start or with business cycle conditions at certain critical ages of teachers. We also show that our results are robust to using alternative business cycle indicators or alternative value-added measures and are not driven by any single recession. Finally, we provide evidence that our results are not driven by differential attrition of recession and non-recession teachers.

4.4.1 Teachers from Recession Entry Cohorts are More Effective in the Classroom

We first present summary statistics separately for recession teachers and the much larger group of non-recession teachers (Table 4.1). The unemployment level of college graduates was higher when recession teachers started their careers. Similarly, unemployment was rising for recession teachers, but slightly falling for non-recession teachers. These differences are significant at the one percent level. The share of male teachers is approximately the same in both samples. Among recession teachers, the share of teachers with a Master's or PhD degree is slightly larger and the share of white teachers somewhat smaller. Because recession teachers started around three school years earlier than non-recession teachers on average, recession teachers also have more teaching experience. The two groups teach similar types of students as measured by the share of students who are black and by the share of students eligible for free or reduced-price lunch. Although none of the teacher characteristics differ significantly, recession teachers have on average 0.08 SD higher math value-added and 0.05 SD higher reading value-added than non-recession teachers.

After documenting the raw gap in math value-added between recession and non-recession

Table 4.1: Summary Statistics by Recession Status at Career Start

	Recession	Non-recession	Diff.	p-Value
Unemp. (college)	2.93	2.24	0.69	0.00
Unemp. change (college)	0.91	-0.12	1.03	0.00
Male	0.12	0.13	-0.01	0.46
Master's or PhD	0.41	0.38	0.03	0.28
White	0.71	0.76	-0.05	0.39
Black	0.15	0.14	0.01	0.15
Hispanic	0.12	0.09	0.03	0.48
Experience	11.06	8.67	2.39	0.62
Career start	1993.98	1996.97	-2.99	0.54
Age at career start	31.26	31.47	-0.21	0.79
Year of birth	1962.72	1965.50	-2.78	0.51
% black (school)	0.25	0.24	0.01	0.55
% free/red. lunch (school)	0.57	0.55	0.02	0.44
VAM (math)	0.07	-0.01	0.08	0.05
VAM (reading)	0.04	-0.01	0.05	0.45
Obs.	5'188	27'946		

Notes: Recession status at career start based on NBER business cycle dates. T-tests adjust for clustering of observations by career start year. Unemployment rates of college graduates only available after 1969 (5'176 and 27'414 observations, respectively); VAM (math) only available for 5'172 and 27'769 observations, respectively. Unemployment data is from the Bureau of Labor Statistics. Teacher and student data stems from the Florida Department of Education.

teachers (see also column 1 in Table 4.2), we add several teacher characteristics (Table 4.2). Due to the idiosyncratic distance between recessions and our sample period, experience is a necessary control. We therefore refer to column (2) as our preferred specification. The value-added gap increases to 0.11 SD when dummies for teaching experience are included (column 2).[17] Adding year of birth and age at career start has little effect on the coefficient on the recession indicator (column 3). Further controlling for teacher characteristics such as whether the teacher holds a Master's or PhD degree, and whether the teacher is male or white, also does not affect our coefficient of interest.[18] The specification with all control variables indicates that recession teachers are 0.10 SD more effective in teaching math than non-recession teachers. Since all control variables except experience represent

---

[17]The coefficient on the recession indicator increases because recession teachers are overrepresented among rookie teachers and the first years of teaching experience improve effectiveness the most.

[18]Differences in the placement of recession and non-recession teachers represent another potential mechanism through which recessions could impact productivity (cf. Oyer, 2006). However, controlling for important student characteristics at the school level, such as the share of black students and the share of students eligible for free or reduced-price lunch, does not explain the value-added difference (results not shown).

potential mechanisms rather than confounders, we omit them in all regressions below.

Table 4.2: Recession at Career Start and Teacher Math Effectiveness

| | Dependent variable: VAM in math | | | |
	(1)	(2)	(3)	(4)
Recession	0.081**	0.110***	0.105***	0.100***
	(0.040)	(0.023)	(0.023)	(0.023)
Year of birth			-0.015***	-0.014***
			(0.005)	(0.005)
Age at career start			-0.020***	-0.019***
			(0.005)	(0.004)
Master's or PhD				0.070***
				(0.010)
Male				-0.037**
				(0.018)
White				-0.053**
				(0.026)
Experience dummies	no	yes	yes	yes
Clusters (career start years)	60	60	60	60
Obs. (teachers)	32941	32941	32941	32941
R^2	0.001	0.022	0.024	0.026

Notes: Regressions of VAM in math on NBER recession indicator at career start. Experience controls include yearly experience dummies up to 30 years. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%.

The simple Roy model predicts selection effects due to changing outside labor-market options over the business cycle. Because research indicates that earnings returns are twice as large for numeracy than for literacy skills in the U.S. labor market (Hanushek et al., 2015), we expect selection effects over the business cycle to be weaker for reading effectiveness than for math effectiveness. The effects on teachers' reading value-added are indeed similar to, but weaker than in math (Table 4.3). However, these results are also consistent with the common finding that students' reading scores are more difficult to improve than their math scores (Jackson et al., 2014, cf.).The bivariate relationship between recession at career start and teacher effectiveness is positive, but statistically insignificant (column 1). As in math, controlling for teaching experience increases the coefficient on the recession indicator; the estimate also becomes significant at the one percent level (column 2). Adding the other teacher characteristics reduces the coefficient of interest only slightly. In terms of magnitude, the recession indicator for reading is half as large as the coefficient for math (around 0.05 SD). As selection effects among

potential teachers should be stronger with respect to math skills, we focus on teachers' math effectiveness in the remaining analyses.[19]

Table 4.3: Recession at Career Start and Teacher Reading Effectiveness

| | Dependent variable: VAM in reading | | | |
	(1)	(2)	(3)	(4)
Recession	0.048	0.051***	0.047***	0.044***
	(0.064)	(0.016)	(0.014)	(0.014)
Year of birth			-0.010**	-0.010**
			(0.004)	(0.004)
Age at career start			-0.012***	-0.012***
			(0.004)	(0.004)
Master's or PhD				0.040***
				(0.013)
Male				-0.139***
				(0.018)
White				-0.027
				(0.019)
Experience dummies	no	yes	yes	yes
Clusters (career start years)	60	60	60	60
Obs. (teachers)	33134	33134	33134	33134
R^2	0.000	0.026	0.027	0.030

Notes: Regressions of VAM in reading on NBER recession indicator at career start. Experience controls include yearly experience dummies up to 30 years. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

While Table 4.2 indicates that recession teachers are on average more effective in raising students' math test scores than non-recession teachers, it is unclear whether this effect is driven by the presence of fewer ineffective teachers or more highly effective teachers in recession cohorts. To analyze the recession impact across the distribution of math value-added, we estimate kernel density plots and quantile regressions. The kernel density plots of teachers' (experience-adjusted) math value-added reveal a clear rightward shift in the math value-added distribution for recession cohorts (Figure 4.2).[20] In quantile regressions that control for experience, we analyze this finding further (Figure 4.3 and Table D.1 in appendix D). While teachers at the very low tail of the value-added distribution have very similar VAMs, recession teachers are more effective than non-recession teachers from the

---

[19]The results of the following analyses show the same overall pattern for teachers' reading effectiveness, but are less pronounced and more volatile than the results for math. All results are available on request.

[20]Kolmogorov-Smirnov tests indicate that the distributions are statistically significantly different at the one percent level.

Figure 4.2: Recession at Career Start and Teacher Math Effectiveness
(Kernel Density Estimates)
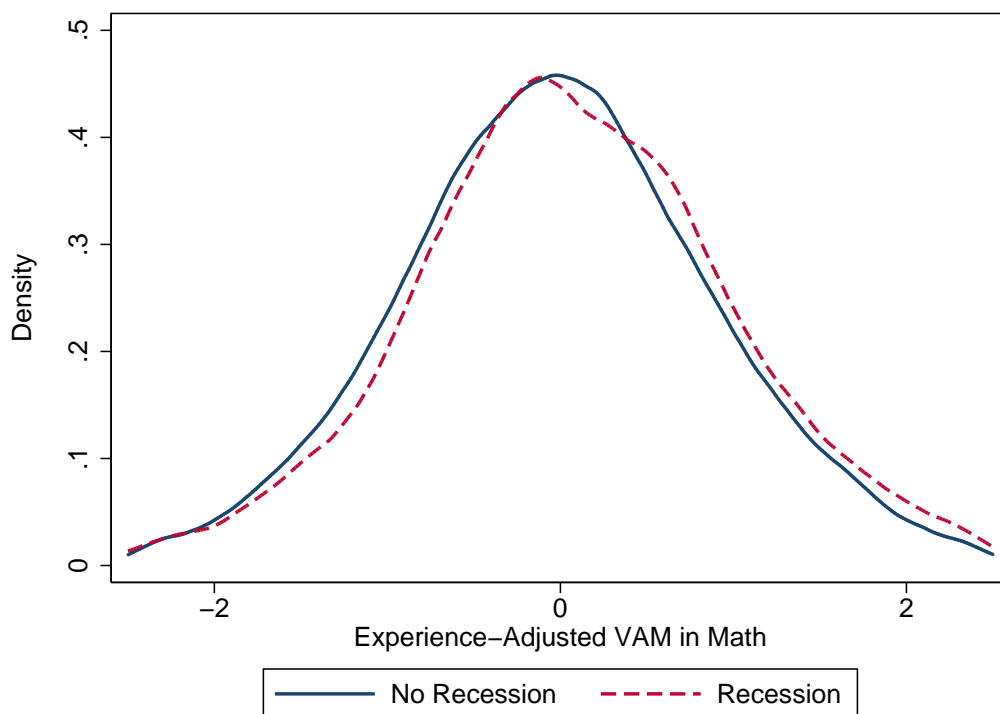
Notes: Kernel density estimates of VAM in math (controlling for yearly experience dummies up to 30 years), by recession cohort status. Excludes teachers with experience-adjusted $|VAM| > 2.5$ for better visibility (805 of 32'941 teachers dropped). VAMs normalized to have mean 0 and standard deviation 1 among all teachers. A Kolmogorov-Smirnov-test shows the distributions are statistically significantly different ($p < 0.01$).

10th percentile onwards. The largest difference between the distributions appears among highly effective teachers, with point estimates of differences peaking at 0.20 SD in the upper end of the distribution.

In Table 4.4, we run our preferred specification on subsamples to assess whether recessions have differential impacts across various groups of teachers. Male teachers seem to be more affected than female teachers (columns 1 and 2) which may suggest that the career options of men are more strongly influenced by recessions than those of women. In columns (3) and (4), we find similar recession impacts for teachers with and without a Master's or PhD degree. In line with existing research (Jones and Schmitt, 2014; Hoynes et al., 2012), columns (5) and (6) provide indirect evidence that minorities are more affected by recessions than whites. Finally, columns (7) and (8) indicate that teachers starting their teaching careers at a relatively high age (above median) are more affected than those

Figure 4.3: Recession at Career Start and Teacher Math Effectiveness (Quantile Regressions)
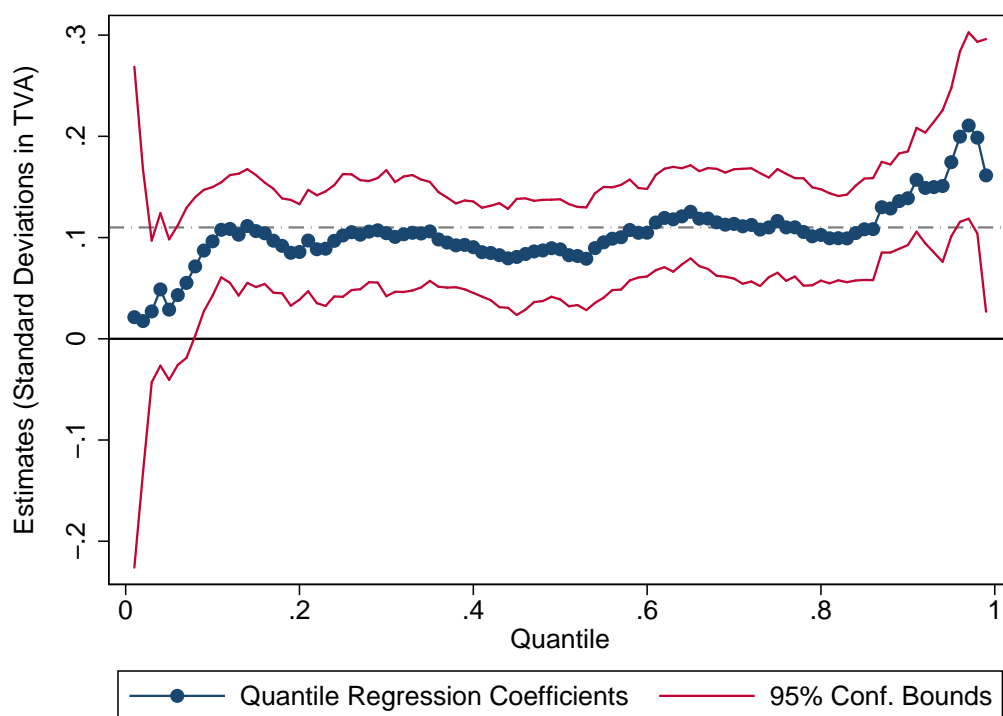
Notes: Coefficients (and 95% confidence bounds) from separate quantile regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on NBER recession indicator at career start at different quantiles. Dashed grey line: OLS estimate from Table 4.2, column (2). Teacher and student data stems from the Florida Department of Education. Standard errors adjusted for clustering at the career start year level.

starting at younger ages. This may suggest that the decisions of mid-career entrants to the teaching profession are more strongly influenced by the outside labor market.

Table 4.4: Recession at Career Start and Teacher Math Effectiveness (Subgroups)

			Dependent variable: VAM in math					
Subsample:	Male	Female	Master's/PhD	Bachelor's	White	Non-white	≤ Median age at career start	>Median age at career start
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Recession	0.164***	0.101***	0.106***	0.105***	0.074***	0.159***	0.091***	0.136***
	(0.039)	(0.021)	(0.029)	(0.029)	(0.025)	(0.039)	(0.022)	(0.040)
Clusters (career start years)	54	60	58	58	58	58	60	45
Obs. (teachers)	4171	28770	12596	20345	24681	8260	17535	15406
R^2	0.033	0.021	0.012	0.027	0.028	0.022	0.023	0.022

Notes: Coefficients from separate regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on NBER recession indicator at career start for different subsamples. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Median age at career start is 29. Significance levels: *** $p < 1\%$, ** $p < 5\%$, * $p < 10\%$

4.4.2   Placebo Analyses Support the Identification Assumption

We assume that it is the business cycle condition at the point in time when individuals enter the teaching profession that matters for their effectiveness. If this is true, then the economic conditions several years before or after career start should be irrelevant. To test this hypothesis, we run placebo regressions where we include recession indicators for the years before or after career start with lags and leads of up to three years. Adding these recession indicators to the main model does not change our coefficient of interest (columns 2 and 3 in Table 4.5). Furthermore, the estimated effects of the business cycle conditions in the years before or after our preferred year are all close to zero and statistically insignificant.[21]

One might worry that our career start year measure captures the effect of macroeconomic conditions at key ages (Giuliano and Spilimbergo, 2014). For example, many individuals may decide to become teachers when entering college (around age 18) or upon completing their undergraduate or graduate studies (between ages 22 to 24). Therefore, we include recession indicators at ages 18-32 (in two-year steps) to confirm that it is the economic conditions at career start that affect teaching quality. As before, all coefficients on the indicators of recessions at specific ages are close to zero and statistically insignificant (column 4).

4.4.3   Further Robustness Checks

Since the number of recession cohorts is limited, one might worry that our result is driven by only one or two recessions. To investigate this issue, we include a separate binary indicator for each recession (Table 4.6).[22] Column (1) indicates that teachers in most recessions (except in recession years 1974; and 1980–82, a highly atypical recession as the demand for teachers decreased, see Figure 4.1) have higher math value-added than the average non-recession teacher. In column (2), we combine the separate recession indicators for the adjacent recession years of 1980, 1981, and 1982 and find that teachers who started

---

[21]Similarly, using each of these other recession indicators individually instead of our main recession indicator also yields small and mostly statistically insignificant coefficients.

[22]Because there are fewer than 20 teachers per cohort who started teaching before 1962, we exclude these cohorts for this analysis since estimates are less reliable for very small cohorts.

Table 4.5: Placebo Analyses: Recession at Different Points in Life and Teacher Math Effectiveness

Recession at:	Dependent variable: VAM in math			
	(1)	(2)	(3)	(4)
Career start	0.110***	0.110***	0.101***	0.104***
	(0.023)	(0.024)	(0.023)	(0.022)
Career start -1 yr.		0.009		
		(0.029)		
Career start -2 yrs.		-0.006		
		(0.020)		
Career start -3 yrs.		0.003		
		(0.025)		
Career start +1 yr.			0.035	
			(0.022)	
Career start +2 yrs.			-0.011	
			(0.021)	
Career start +3 yrs.			-0.028	
			(0.026)	
Age 18				-0.006
				(0.015)
Age 20				0.007
				(0.018)
Age 22				-0.016
				(0.012)
Age 24				-0.017
				(0.015)
Age 26				-0.022
				(0.014)
Age 28				-0.025
				(0.017)
Age 30				-0.026
				(0.017)
Age 32				0.011
				(0.018)
Clusters (career start years)	60	60	60	60
Obs. (teachers)	32941	32941	32941	30038
R^2	0.022	0.022	0.022	0.020

Notes: Regressions of teacher VAM in math on NBER recession indicator (controlling for yearly experience dummies up to 30 years) at different points in time. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** $p< 1\%$, ** $p< 5\%$, * $p< 10\%$

Table 4.6: Recession at Career Start and Teacher Math Effectiveness (Single Recessions)

	Dependent variable: VAM in math		
Recession year (career start)	(1)	(2)	(3)
1970	0.102***	0.102***	0.080**
	(0.029)	(0.029)	(0.036)
1974	0.020	0.020	0.009
	(0.020)	(0.020)	(0.025)
1980	0.017	-0.004	-0.034
	(0.035)	(0.034)	(0.034)
1981	0.002		
	(0.033)		
1982	-0.034		
	(0.031)		
1990	0.076***	0.076***	0.092***
	(0.016)	(0.016)	(0.009)
2001	0.138***	0.138***	0.124***
	(0.016)	(0.016)	(0.023)
2008	0.264***	0.264***	0.230***
	(0.036)	(0.036)	(0.049)
Included cohorts:			+/- 2 years
	all	all	around recessions
Clusters (career start years)	48	48	28
Obs. (teachers)	32897	32897	19144
R^2	0.023	0.023	0.023

Notes: Regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on seperate dummies for cohorts starting during each NBER recession (recession cohorts). Excludes observations with fewer than 20 teachers; mean teacher cohort size is 1'292. In columns (2) and (3), cohorts entering in 1980 through 1982 are combined. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

during those years are on average as effective as the average non-recession teacher. In column (3), we only keep two non-recession cohorts immediately before and immediately after each recession, such that the cohorts being compared are more similar. This leads to the same finding: most recessions have positive effects on teacher effectiveness. The recession impact is not driven by any single recession.

Table 4.7: Recession at Career Start and Teacher Math Effectiveness (Subsamples)

	Dependent variable: VAM in math				
	(1)	(2)	(3)	(4)	(5)
Recession	0.110***	0.018	0.149***	0.185***	0.140***
	(0.023)	(0.025)	(0.025)	(0.033)	(0.022)
Subsample:	Baseline	Pre-1990 cohorts	Post-1990 cohorts	Full teacher career observed	No exp. in other state
Clusters (Car. start yrs.)	60	40	20	9	58
Obs. (Teachers)	32941	7303	25638	15731	26709
R^2	0.022	0.003	0.025	0.028	0.028

Notes: Regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on recession indicator at career start. In column (1), the recession indicator only takes the value of one for the recessions before 1990. In column (2), the recession indicator takes the value of one for the recessions since 1990, including 1990. Columns (3) and (4) use our preferred recession indicator and use subsamples as indicated in the text. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

In our main analyses, we use the variation in business cycles across teacher cohorts that started their careers many years before our sample period begins. To assess whether recent recessions matter more for current teacher quality than distant recessions, in Table 4.7 we present estimates of the impact of a recession at career start on teacher value-added separately for recent and distant teacher cohorts. Columns (2) and (3) show that the impact of recent recessions is higher than the baseline estimate and that the impact of distant recessions is small and not significant. This could reflect differences in the returns to experience or differential patterns of attrition with respect to effectiveness among recession and non-recession teachers, an issue we examine directly in Section 4.4.4.

Since we estimate the year of career start, we cannot observe gaps in teachers' careers due to fertility, child-rearing or family mobility before our sample period begins. To assess whether our results are sensitive to this, column (4) restricts the sample to the entry cohorts for which we can observe the entire career. The estimate is larger than the baseline effect and significant at the one percent level. However, because this very short

panel only contains two recessions, we prefer to use all available entry cohorts. Finally, we test whether our estimates reflect selection into the teaching profession or selection of teachers with experience elsewhere into Florida public schools. In column (5), we restrict the sample to those teachers without any teaching experience outside Florida. The coefficient is somewhat larger than in the baseline specification.[23]

Figure 4.4: One-Year Unemployment Change and Mean Teacher Math Effectiveness

Notes: Cohort means of VAM in math (controlling for yearly experience dummies up to 30 years) and one-year unemployment change for college graduates. Teacher and student data stems from the Florida Department of Education. Unemployment rates come from the BLS. 2008-09 cohort excluded as outlier (unemployment change=2.2, mean experience-adjusted VAM=0.21).

We also evaluate the robustness of our results using alternative measures of teachers' outside options. Figure 4.4 makes it possible to compare the variation in our preferred binary measure of the business cycle (by comparing green and blue dots) and a continuous measure, one-year unemployment changes. In line with our main findings, unemployment changes and teacher value-added are positively related. Figure 4.5 displays the variation

---

[23]Moreover, there is no statistically significant difference in the the incidence of teaching experience outside Florida between recession (20.9%) and non-recession cohorts (18.5%). Controlling for any out-of-state experience does not change our coefficient of interest either. This makes an explanation based on migration patterns into Florida unlikely.

of both our value-added measure and the one-year unemployment change over time. The time series move very closely, especially in the more reliable sample of teachers who started their careers after 1990. In Table 4.8, we run our preferred specification using the NBER recession indicator (column 1), GDP growth (2), the unemployment level (3), and one-year unemployment changes (4), respectively. Both unemployment measures are computed using the unemployment rates of college graduates (only available from 1970 onwards), as this is the relevant labor market for potential teachers.[24]

Table 4.8: Recession at Career Start and Teacher Math Effectiveness (Alternative Business Cycle Measures)

| | Dependent variable: VAM in math | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Recession	0.110***					
	(0.023)					
GDP growth		-0.014**				
		(0.006)				
Unemp. (college)			0.052**			
			(0.022)			
Unemp. change (college)				0.083***		
				(0.015)		
Nonagriculture industries					0.040***	
					(0.011)	
Agriculture industries						0.015
						(0.010)
Clusters (career start years)	60	60	40	39	57	57
Obs. (teachers)	32941	32941	32402	32244	32936	32936
R^2	0.022	0.021	0.021	0.022	0.022	0.021

Notes: Coefficients from separate regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on alternative business cycle measures at career start. Unemployment (college) refers to BLS unemployment rates of college graduates (4 years and above until 1991, degree holders after 1991) and are available after 1969. All unemployment rates are from the BLS; GDP growth (2009 constant dollars) from the BEA. Agriculture industries refers to private wage and salary workers. Standard errors in parentheses adjusted for clustering at the career start year level. Teacher and student data stems from the Florida Department of Education. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

Consistent with our preferred business cycle indicator, GDP growth is negatively related to teacher value-added. The coefficients on the unemployment measures are also in line with

---

[24]The results of our preferred specification are unchanged for teachers starting after 1970. We use national rather than Florida-specific unemployment rates in this analysis because state-level unemployment rates are not available for college graduates, the national unemployment rates are more reliable, and because Florida recruited teachers heavily from out of state throughout our sample period (see Section 4.3.1). Thus, using Florida-specific measures of economic conditions is likely to underestimate the true effect. In Table D.3 in appendix D, we show that graduate-specific unemployment rates have a stronger impact on teacher value-added than general national unemployment rates and that Florida-specific unemployment rates have around the same impact than national unemployment rates.

our previous findings and significant at the five percent level. The coefficient estimates for the alternative measures imply somewhat weaker, but qualitatively similar recession effects (based on the difference in each business cycle indicator between recession and non-recession cohorts), suggesting that none of the alternative business cycle indicators on its own fully captures the full effects of a recession on potential teachers' choices.[25] Finally, it is unlikely that the alternative job opportunities of potential teachers are evenly distributed across industries. For example, one would expect few potential teachers to work in agriculture. In columns (5) and (6), we find that the one-year unemployment change in agriculture at career start is unrelated to teacher quality, while the labor-market conditions in nonagriculture industries do matter. This pattern is consistent with the selection of potential teachers into teaching who alternatively would have chosen industries requiring similar skills.

To assess the sensitivity of our results with respect to the value-added measure, we also run our preferred specification with alternative VAMs (Table D.2 in the appendix). For comparison, column (1) presents the results based on our preferred measure. In column (2), we add school fixed effects when estimating teachers' value-added. The inclusion of school fixed effects eliminates any bias from unobserved school characteristics that influence teacher effectiveness, but also removes variation in true teacher effectiveness to the extent that average teacher quality varies across schools. The gap in effectiveness between recession and non-recession teachers is somewhat attenuated, but the change is small. In column (3), we add school-by-year fixed effects when estimating value-added, likely removing additional variation in true teacher effectiveness. The estimate is further attenuated, but remains significant. Finally, in columns (4) and (5), we account for the fact that the precision of the teacher value-added measures varies across teachers. Our results are qualitatively unaffected by weighting teachers in our preferred specification by the number of student-year or teacher-year observations that underlie their value-added measures.

---

[25]The same pattern appears if we use unemployment rates and changes for all workers rather than college graduates. These coefficients are significant at the one percent level, but somewhat attenuated, as expected.

Figure 4.5: One-Year Unemployment Change and Mean Teacher Math Effectiveness over Time
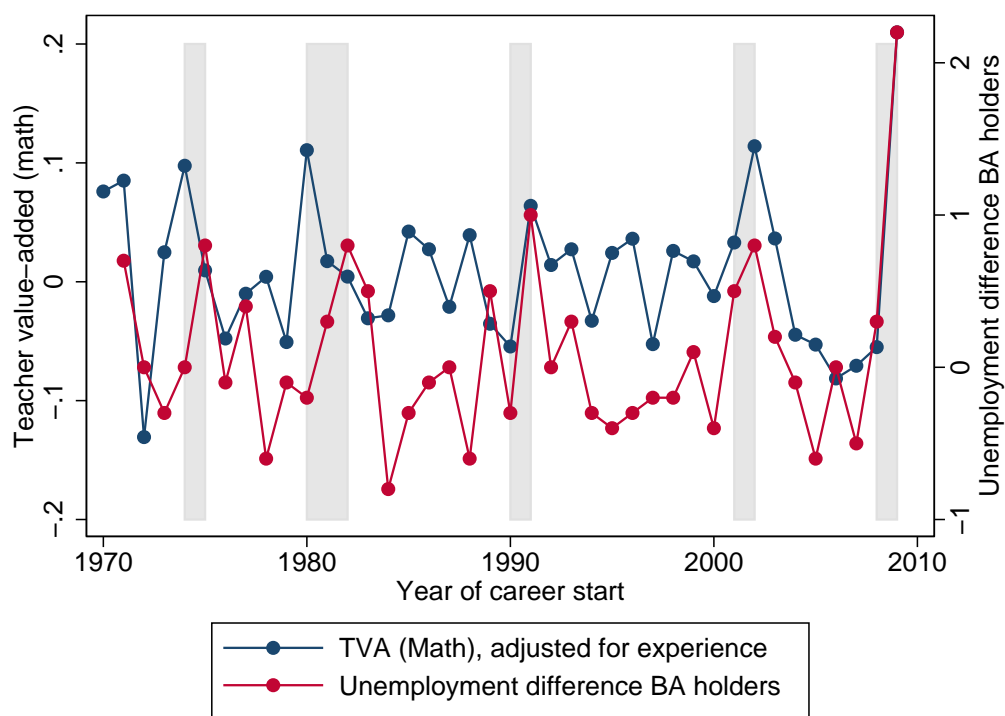
Notes: Cohort means of VAM in math (controlling for yearly experience dummies up to 30 years) and one-year unemployment change for college graduates. Unemployment rates from the BLS. Teacher and student data stems from the Florida Department of Education. Shaded areas are recession periods as defined by the NBER.

4.4.4 Differential Attrition of Teachers does not Drive Results

We find that teachers who started their careers during recessions are more effective. On the one hand, effectiveness differences might already exist among entering teachers (*selection*). On the other hand, recession and non-recession teachers might have very similar VAMs at career start, but low-quality recession teachers might be more likely to leave the occupation than low-quality non-recession teachers (*differential attrition*). We use our data to assess which of these two channels is more plausible.

Since our dataset includes all teachers in the public school system in Florida, attrition means that a teacher leaves the Florida public school system. We cannot directly address attrition before 2000-01, the beginning of our sample period. However, if differential attrition of recession and non-recession teachers were driving our results, then one would

expect earlier recession cohorts to be much more effective, but more recent recession cohorts to be only slightly more effective, than non-recession teachers. This pattern is not present in Table 4.7, which shows that recession effects are generally larger for more recent cohorts. We interpret this as first, indirect evidence that differential attrition does not drive our results.

To provide direct evidence, we define attrition as *not* being observed as a teacher during the last school year in our sample period (2008-09). First, we investigate whether starting during a recession is correlated with attrition (columns 1 and 2 in Table 4.9).[26] Controlling for teachers' value-added, we find that recession teachers are somewhat more likely to drop out, although this difference is not statistically significant. Controlling for recession status at career start, more effective teachers are less likely to drop out.[27]

Table 4.9: Recession at Career Start, Attrition,
and Teacher Math Effectiveness

Dependent variable:	Attrition		VAM in math	
	(1)	(2)	(3)	(4)
Recession	0.039	0.017	0.182***	0.333***
	(0.039)	(0.029)	(0.026)	(0.033)
VAM (math)	-0.029***	-0.048***		
	(0.005)	(0.009)		
Recession*VAM (math)	0.005	0.039***		
	(0.012)	(0.009)		
Career start	-0.004***	-0.040***		
	(0.001)	(0.010)		
Recession*experience			-0.007***	-0.074***
			(0.002)	(0.010)
Included cohorts:	<2008	2000-07	all	2000-08
Clusters (career start years)	59	8	60	9
Obs. (teachers)	32417	15207	32941	15731
R^2	0.013	0.043	0.023	0.031

Notes: Regressions of attrition indicator (columns 1 and 2) and VAM in math (columns 3 and 4) on regressors as shown in table. Attrition defined as no teacher observation in 2009. Columns (3) and (4) control for yearly experience dummies up to 30 years. Teacher and student data stems from the Florida Department of Education, the recession indicators are coded as defined by the NBER. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

Among teachers who started teaching during our sample period (about 47% of the full

---

[26]Because the school year 2008-09 is the attrition target year, these regressions exclude teachers who started teaching in 2008-09.

[27]Excluding teachers born before 1950 as potential retirees does not change our results (not shown).

sample), recession teachers are also slightly more likely to leave the public school system than non-recession teachers (column 2). More importantly, in recession cohorts, exiting teachers are significantly more effective compared to exiting non-recession teachers. This pattern works against our result, suggesting that the value-added gap is even larger at career start and decreases over time. This is confirmed in column (3) when we look directly at value-added, finding a large gap at career start which decreases with experience. Taken at face value, these estimates imply that the gap in value-added between recession and non-recession teachers closes after around 25 years. However, depending on the functional form we impose on the interaction between starting in a recession and teaching experience, the implied time period before the gap closes ranges from 12 to 26 years. Therefore, these numbers need to be interpreted very cautiously. Column (4) confirms that the same pattern holds, and in fact becomes more pronounced, when using only teachers who started teaching during our sample period.

In sum, differential attrition between recession and non-recession teachers does not explain our main finding. The observed attrition pattern seems to reduce the estimated difference in effectiveness between recession and non-recession teachers over time. This suggests that our main results understate the difference in effectiveness between recession and non-recession teachers at career start.

4.4.5   Discussion

The effect of recessions at career start on teacher effectiveness might in theory be driven by demand or supply fluctuations over the business cycle (or both). As noted in Section 4.2, demand fluctuations can generate our findings only if school authorities (i) hire fewer teachers during recessions (e.g., due to budget cuts) *and* (ii) are able to assess the quality of inexperienced applicants and hire those most likely to be effective. Both conditions are unlikely to hold in practice. First, in our data, cohort size is unrelated to the business cycle. This is corroborated by official statistics from the BLS, which indicate that employment in the local government education sector typically increases during recessions (with the exception of the recessions in 1980-1982 and the Great Recession; see Figure 4.1 and Berman and Pfleeger, 1997). Second, it is unlikely that school authorities

are able to identify the best applicants since education credentials, SAT scores, and demographic characteristics – typically the only ability signals of applicants without prior teaching experience – are at best weakly related to teacher effectiveness as measured by VAMs (e.g., Chingos and Peterson, 2011; Jackson et al., 2014). Apart from the fact that both conditions are unlikely to hold, our quantile regression results show that the effect is strongest at the upper end of the value-added distribution. This suggests that increases in the supply of very effective teachers rather than decreases in the overall demand for teachers are at work.[28]

In sum, increases in the supply of high-quality applicants during recessions seem to drive our results. Teacher cohorts likely differ in their effectiveness already at career start, as predicted by a Roy model of occupational selection.

Finally, note that we estimate a reduced-form coefficient. To gauge the quality difference between recession-only teachers and those they replace, we have to inflate our reduced-form estimates by the share of recession-cohort teachers who would not have entered teaching under normal labor-market conditions. If all teachers who start during recessions became teachers only because of the recession, the effectiveness difference would be equal to our reduced-form estimate (0.11 SD). However, if only 10% of the recession teachers went into teaching due to the recession, the difference in effectiveness would be 10 times as large, around one SD. This would imply an impact on student math achievement of being assigned to a recession-only entrant of around 0.2 student-level standard deviations.

4.5   Policy Implications

Our results have important implications for policymakers. In a Roy model of occupational choice, worse outside options during recessions are equivalent to higher teacher wages.

---

[28]In emphasizing the role of high-quality supply, we further assume that recessions have no direct effects on teachers' effectiveness. This would be violated, for example, if teachers who started their career in a recession were more fearful of losing their jobs and thus provided more effort, which raised their effectiveness permanently. However, in this case we would expect the least effective teachers to disproportionally better in recession cohorts. In our quantile regressions, we find that the opposite is true. If the business cycle at career start did have a direct effect on the individual's teaching effectiveness, we would estimate the total effect of starting in a recession on subsequent career productivity in teaching, comprising the combined effect of selection into teaching and the direct impact on individual's productivity in teaching. The reduced-form estimate still represents a causal effect.

Thus, our results suggest that policymakers would be able to hire better teachers if they increased teacher pay. Would such a policy be efficient? Chetty et al. (2014b) find that students taught by a teacher with a one SD higher value-added measure at age 12 earn on average 1.3% more at age 28. Using this figure, our preferred recession effect translates into differences in discounted lifetime earnings of around $13,000 per classroom taught each school year by recession and non-recession teachers (evaluated at the average classroom size in our sample). This is equivalent to more than 20% of the average teacher salary in Florida ($46'583 in school year 2012-2013 according to the Florida Department of Education).

Do these private benefits exceed the public costs associated with an increase in teacher pay intended to attract more effective teachers? To shed light on this question, assume that the entire recession effect is driven by earnings losses in the private sector during recessions. To compute these earnings losses, we use the median earnings of BA degree holders ($59'488 in 2010, the year Chetty et al.'s figure refer to) as a benchmark for the average outside option of potential teachers. The adverse impact of graduating in a recession has previously been estimated to be around 2%–6% of initial earnings per percentage point increase in the unemployment rate (e.g., Kahn, 2010). This translates into 4%–12% earnings differences between recession and non-recession teachers in our sample. Based on the median earnings of BA degree holders, this implies on average between $2'379 and $7'140 lower earnings during recessions. This admittedly coarse comparison suggests that it may be efficient to increase pay for new teachers and thereby improve average teacher effectiveness. Yet this conclusion comes with the caveat that it may be difficult for policymakers to increase pay only for incoming teachers. Our evidence does not imply that increasing pay for the existing stock of teachers would yield benefits. Moreover, there are likely cost-neutral ways to make the total compensation package offered to new teachers more attractive. For example, Fitzpatrick (2015) shows that the value teachers place on pension benefits is much lower than the cost to the government of providing them and would prefer higher salary levels.

Magnitudes aside, our findings suggest that policymakers would be able to attract more effective individuals into the teaching profession by raising the economic benefits of becoming a teacher. This is not a trivial result. If intrinsic motivation positively affects teachers'

effectiveness, then increasing teacher pay may attract more extrinsically motivated, but less effective individuals into the teaching profession. Since we find the opposite, intrinsic motivation seems to be of second-order importance relative to the effects of increasing teacher pay on selection when hiring more effective teachers.

Finally, our results indicate that recessions serve as a window of opportunity for the public sector to hire more effective personnel than during normal economic periods. As teachers are a critical input in the education production function affecting students' lives way beyond schooling, hiring more teachers in economic downturns would appear an attractive strategy to improve American education. In the Great Recession, however, even substantial stimulus spending was insufficient to prevent a reduction in employment in the education sector (see Figure 4.1).

4.6   Conclusion

We provide causal evidence on the importance of outside labor-market options at career start for the quality of teachers. We combine a novel identification strategy with a direct and well-validated measure of teacher effectiveness. Our reduced-form estimates show that teachers who entered the profession during recessions are significantly more effective than teachers who entered the profession during non-recessionary periods. This finding is best explained by a Roy-style model in which more able individuals prefer teaching over other professions during recessions due to less opportunities in alternative occupations. In comparison to Britton and Propper (2016), we show that the *selection into teaching* is affected by outside options. We can additionally control for potential confounding channels by using individual-level data and a direct measure of teacher quality. While the settings differ, our productivity effects are qualitatively similar to, and in fact somewhat larger than, recession effects on the productivity of PhD economists (Boehm and Watzinger, 2015). Recessions may serve as a window of opportunity for recruitment in the public sector.

Appendices

Appendix A

Appendix to Chapter 1

Compulsorily Licensed Patents by NBER Technological Subcategory

Figure A.1: Compulsorily Licensed Patents by NBER Technological Subcategory
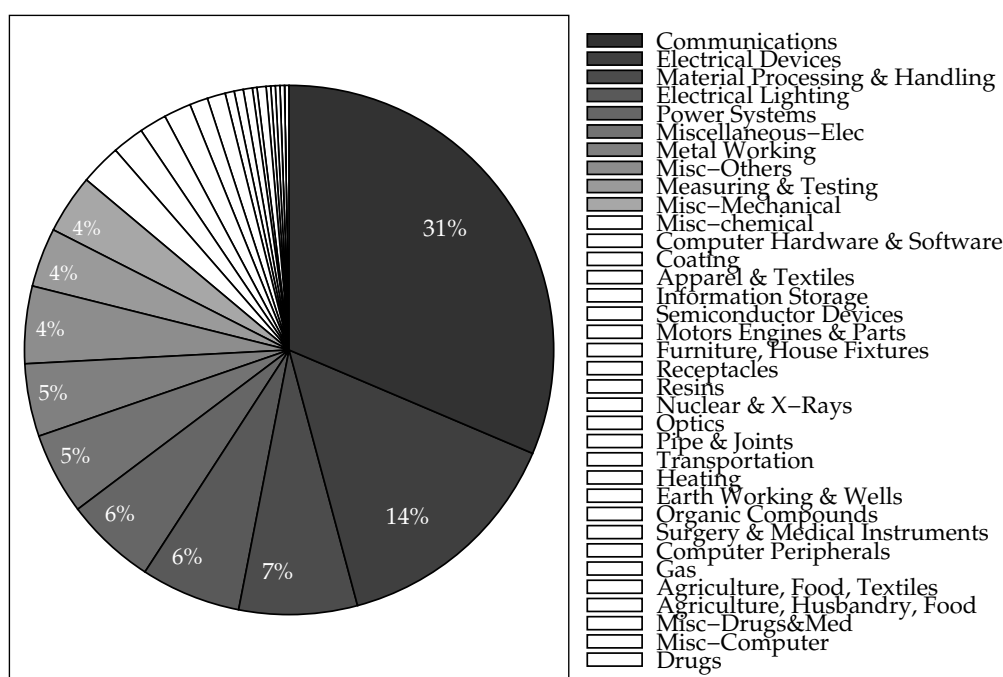
Notes: The pie chart shows the distribution of compulsorily licensed patents over 35 NBER technological subcategories. The legend is sorted from largest share to smallest. The categorization in technological subcategories is based on US patent classifications, following Hall et al. (2001). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Patenting of Bell in Radar and Cryptography

Figure A.2: War Technologies Created by Bell Labs
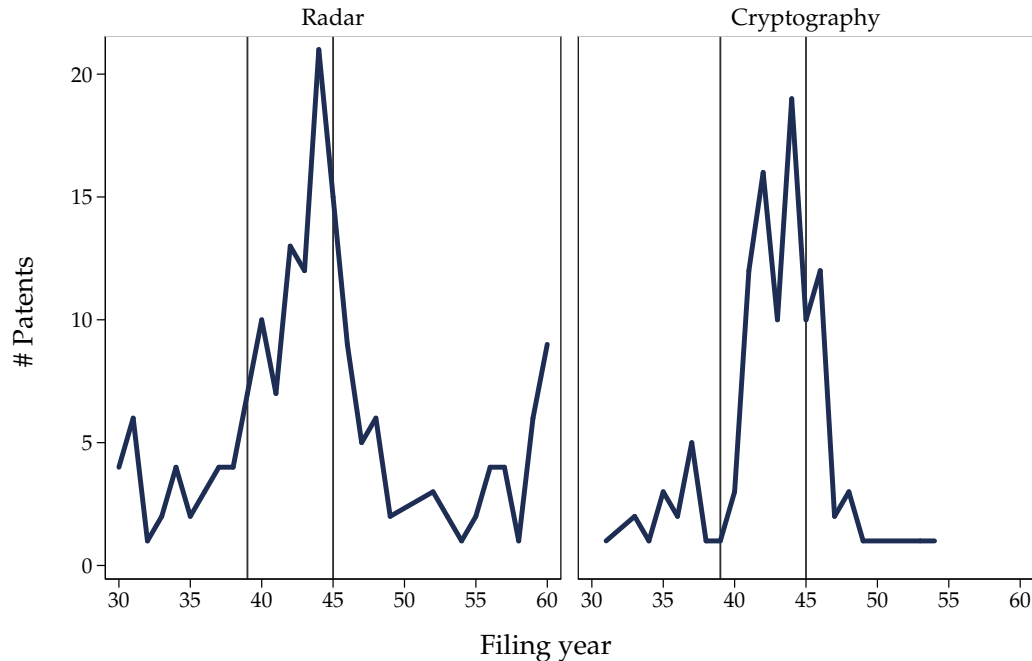
Notes: This figure shows the yearly number of Bell patents relating to radar and cryptography, two technologies relevant for World War II. We identify both technologies by their USPC class: We use the class 342 titled "Communications: directive radio wave systems and devices (e.g., radar, radio navigation)" to classify radar and class 380 titled "Cryptography" to classify cryptography. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

A.1 Appendix to Section 1.4

Comparing the Average Number of Citations of Treatment and Control Patents

In Figure A.3 we compare the evolution of patent citations to Bell patents and control patents in the same publication year and the same four digit technology class. We use the weights proposed of Iacus et al. (2009) to adjust for the different number of control patents for each Bell patent. From 1949 to 1953, the average number of citations of treatment

Figure A.3: Average Number of Citations to Bell and Control Patents Published before 1949
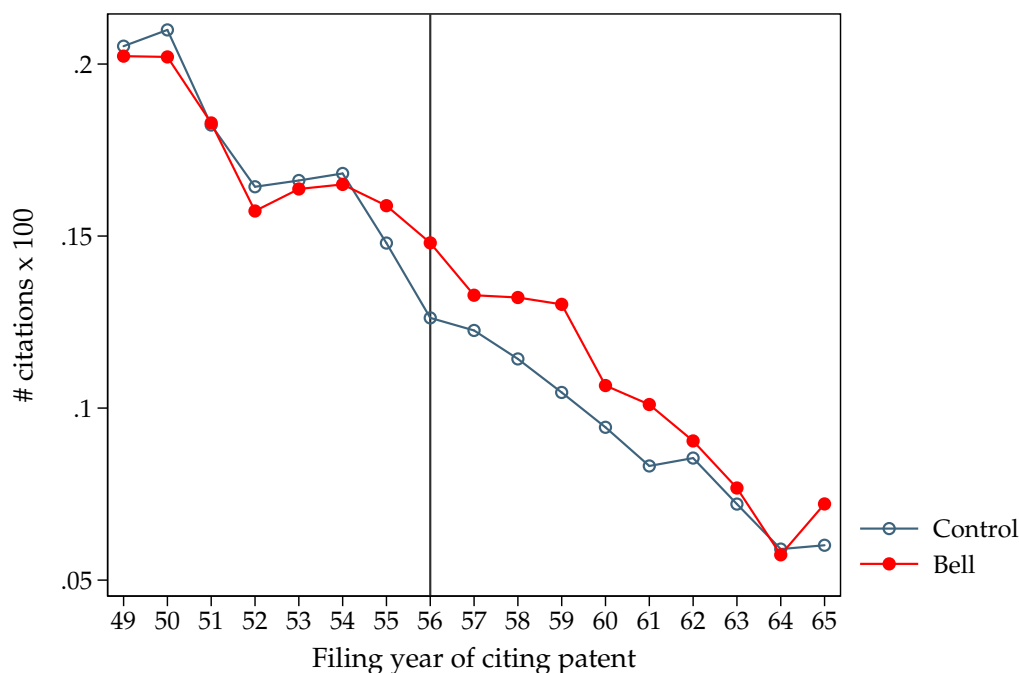
Notes: This figure shows average patent citations of patents published before 1949 in every year after publication. The line with solid circles shows patent citations of the treated patents (Bell patents) and the line with empty circles shows patent citations of control patents, with the same publication year and the same four digit technology class as the Bell patents. For aggregation we use the weights of Iacus et al. (2009) to adjust for a different number of control patents for each Bell patent. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

and control patents track each other very closely. This implies that the Bell patents and the control patents exhibit a parallel trend in citations in the first 4 years after the plea. The two lines diverge in 1954, with Bell patents receiving relatively more citations than control patents, and they converge again in 1961/1962. This is prima facie evidence for an effect from 1954 onward.

Pseudo Outcomes: Unaffected Companies have no Excess Citations

In the main part of the text we use time varying coefficients to show that there are no yearly excess citations from the B-2 companies, which were exempt from the compulsory

licensing agreement. In column (2) of Table A.1 we estimate the average effect for these companies and find none. There are also two other groups of companies that were to a lesser degree affected by the consent decree: foreign companies and companies that already had licensing agreements in place.[1] Foreign companies could license for free but did not receive any technical description or assistance from Bell.[2] In Table A.1 we show the results using as the dependent variable the citations from from foreign companies in column (3) and from companies that had a license before the consent decree in column (4). In the last column we use data on all companies that did not have a license from Bell. We do not find a measurable effect for foreign companies or companies with a license and a large effect for companies without a license.

Pseudo Treatment: Citation Substitution is Small.

One possible interpretation of our estimates is that due to the free availability of Bell technology, companies substituted away from other, potentially more expensive technologies. If this were the case, we should find a negative impact of the consent decree on citations of similar patents of other companies.[3] To see if this is the case, we assign a pseudo treatment to the patents of GE, RCA, Westinghouse, which were part of the B-2 agreement, and ITT. These companies were among the largest patenting firms in the ten technology classes in which Bell had most patents between 1939 and 1949. Results are reported in Table A.2, column (2). We find no effect, implying that the citation substitution is either small or homogeneous to patents of these companies and the control group.

---

[1]All companies with a license agreement are listed in the hearing documents (Antitrust Subcommittee, 1958, p. 2758).

[2]Verbatim in the consent decree "The defendants are each ordered and directed (...) to furnish to any person domiciled in the United States and not controlled by foreign interests (...) technical information relating to equipment (...)".

[3]This approach is suggested by Imbens and Rubin (2015).

Table A.1: The Effect of Compulsory Licensing on Subsequent Citations of Unaffected Companies

	(1) Baseline	(2) B-2 Companies	(3) Foreign companies	(4) License	(5) No license
Treatment	-0.4	-0.1	-0.0	0.5***	-0.9**
	(0.5)	(0.2)	(0.1)	(0.2)	(0.4)
I(55-60)	-6.4***	-1.2***	2.1***	-1.1***	-5.4***
	(0.6)	(0.2)	(0.3)	(0.2)	(0.5)
T x I(55-60)	2.0***	0.2	-0.0	0.4	1.6***
	(0.6)	(0.1)	(0.2)	(0.3)	(0.5)
Constant	18.3***	2.3***	0.9***	3.1***	15.2***
	(1.2)	(0.3)	(0.1)	(0.3)	(1.0)
# treated	4533	4598	4533	4533	4533
Clusters	225	225	225	225	225
Obs.	896556	1096212	896556	896556	896556

Notes: This table shows the results from a difference-in-differences estimation with years 1949-1954 as pre-treatment period and 1955-1960 as treatment period. The estimation equation is

$$\#Citations_{i,t} = \beta_1 \cdot Bell_i + \beta_2 \cdot I[1955-1960] + \beta_3 \cdot Bell_i \cdot I[1955-1960] + \varepsilon_{i,t} \qquad (A.1)$$

where $I[1955-1960]$ is an indicator variable for the treatment period 1955-1960. The variable "Bell" is an indicator variable equal to one if a patent is published by a Bell System company before 1949 and therefore treated by the consent decree. As dependent variable we use in the first column all citations by companies other than the filing company. In the second column we use all citations of companies exempt from the consent decree (GE, RCA, Westinghouse & ITT) and in the third column all citations of foreign companies. In the fourth column we use citations of companies that had no licensing agreement with any Bell company prior to the consent decree and in the last column we look at the citation of companies that had a licensing agreement. As control patents, we use all patents that were published in the U.S. matched by publication year, primary United States Patent Classification (USPC) technology class and the number of citations up to 1949. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. All coefficients are multiplied by 100 for better readability. Standard errors are clustered on the three-digit USPC technology class level and *, **, *** denote statistical significance on 10%, 5% and 1% level, respectively.

Table A.2: Auxiliary Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Pseudo Treatment			Diff. Control Group	
	Baseline	B-2 Comp.	Control: Same USPC, diff IPC	Control: Same USPC, diff USPC	Control: Same IPC, diff USPC	Control: Same IPC	Loose
Treatment	-0.4	-0.6	0.7**	-0.1	-0.3	0.4	0.7
	(0.5)	(0.4)	(0.3)	(0.5)	(1.6)	(0.7)	(0.8)
I(55-60)	-6.4***	-4.2***	-1.5***	-2.2***	-6.4***	-6.0***	-6.7***
	(0.6)	(0.4)	(0.5)	(0.2)	(0.7)	(0.5)	(0.6)
T x I(55-60)	2.0***	0.0	-0.5	-0.3	1.2	1.5**	1.5**
	(0.6)	(0.4)	(0.5)	(0.4)	(0.8)	(0.6)	(0.6)
Constant	18.3***	16.6***	11.9***	13.3***	19.7***	17.8***	18.6***
	(1.2)	(0.6)	(0.5)	(0.3)	(1.0)	(1.0)	(1.0)
# treated	4533	7869	42607	48526	4649	4511	4665
Clusters	225	207	202	398	398	386	230
Obs.	896556	836172	760452	1348440	835188	705612	1301412

Notes: This table shows the results from a difference-in-differences estimation with years 1949-1954 as pre-treatment period and 1955-1960 as treatment period, controlling for year fixed effects. As dependent variable, we use all citations by companies other than the filing company. As control patents, we use all patents that were published in the U.S. matched by publication year, primary USPC technology class, and the number of citations up to 1949. In all columns we match the control patents on publication year and the number of citations prior to 1949. In columns (2) to (4) we assign pseudo treatments. In column (2) we define patents of the B-2 companies (GE, RCA, Westinghouse & ITT) as treated and match the control patents on the USPC class. In column (3) we assign all patents that have the same USPC and different 3-digit IPC technology class as treated and in column (4) we assign patents with the same IPC and different USPC classification as treated. In column (5) we use as controls patents in the same IPC 3 class but in a different USPC class than the Bell patents. In column (6) we use as controls patents with the same 4-digit IPC class as the Bell patents. In column (7) we coarsen the publication year to two year windows and sort all pre-citations in 10 equally sized bins to match a larger number of patents. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office. All coefficients are multiplied by 100 for better readability. Standard errors are clustered on the three-digit USPC technology class level and *, **, *** denote statistical significance on 10%, 5% and 1% level, respectively.

For a second approach, we exploit the fact that a patent's technology is classified twice: once in the USPC system, which has a technical focus, and once in the IPC system, which reflects more closely the intended industry or profession ("usage") (Lerner, 1994). In columns (3) and (4) of Table A.2 we assign a pseudo-treatment to all patents that have the same USPC class and the same IPC class as the Bell patents. As control group we use in column (3) patents with the same USPC, but a different IPC classification as Bell patents. In column (4) we use as a control group patents with the same IPC, but a different USPC classification as Bell patents. Thus we compare patents that are arguably more similar to the Bell patents to two different control groups. We find a small, negative but statistically insignificant effect. Again, this speaks in favor of limited citation substitution or - alternatively - a homogeneous citation substitution to all control groups.

Effects are Robust to Different Matching Strategies.

In columns (5) to (7) of Table A.2 and in Figure A.4 we report results from using several alternative matching variables. In the main specification, we use the age (measured by the publication year), the technology (measured by USPC class) and the quality of a patent (measured by the number of citations up to 1949). In column (6) we use patents in the same IPC but different USPC class instead of using those in the same USPC class. In column (7) we match on the IPC classification, independent of the USPC class. Finally, in column (8) we do a coarsened exact matching in order to match all Bell patents.[4] In all three cases the size of the effects is similar to the one in the main specification. In Figure A.4 we show the size of the treatment effects for different combinations of background variables as proxy for age, technology and quality. On the vertical axis we plot the number of matched patents. The coefficient is mostly around 2.

---

[4]Coarsened exact matching was proposed by Iacus et al. (2012). In this specification we match on one of five publication year categories that contain 2 years each and one of 10 prior-citation categories.

Figure A.4: Treatment Effects for Different Matching Variables
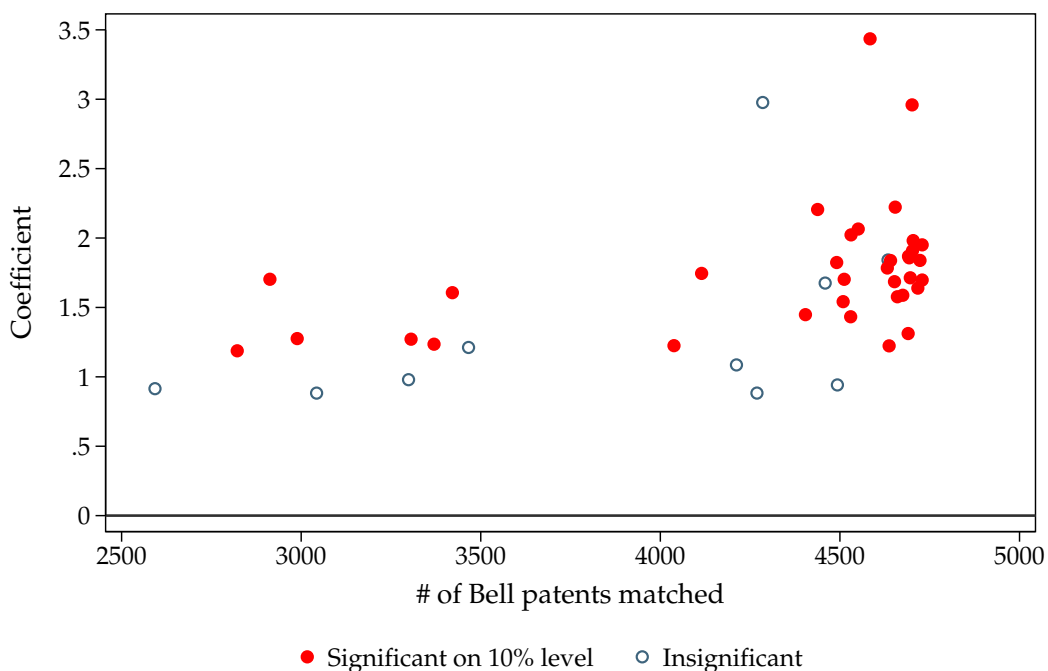
Notes: In this figure we plot the parameter estimates from difference-in-differences estimations of the impact of the consent decree for different matching strategies, controlling for year fixed effects. As before, as dependent variable we use all citations by companies other than the filing company. In all regressions, we use a measure for the age, the technology and the quality of a patent for matching. As measures for the age of a patent, we alternatively use application year, publication year or both. For technology, we use the USPC, the USPC with subclasses, the three and the four digit IPC. As a measure of quality, we use the number of pre-citations as exact numbers, coarsened to steps of five citations and an indicator for at least one citation prior to 1949. The horizontal axis displays the number of matched Bell patents. Empty symbols are insignificant and full symbols are significant at the 10% level. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Patenting Behavior of Bell Relative to Comparable Companies

Figure A.5: Patenting of Bell System and B-2 Companies without RCA
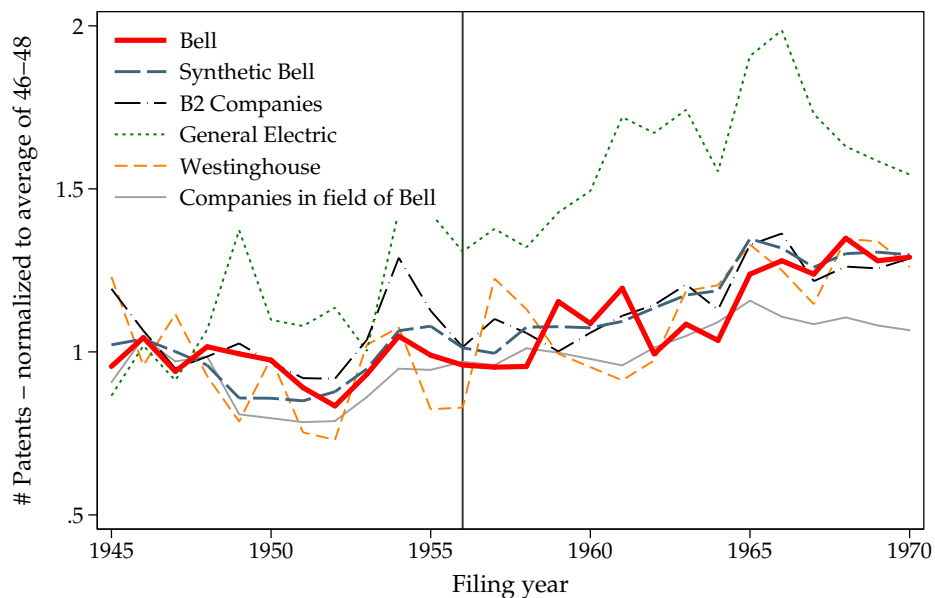
Notes: In this figure we compare Bell's total patenting to a synthetic Bell, the number of patents filed by the B-2 companies (General Electric, Westinghouse, RCA and ITT), General Electric and Westinghouse separately and all companies that existed before 1949 and had at least 100 patents in any field in which Bell was active. The number of patents are normalized to the average number of patents from 1946-1948. We show General Electric and Westinghouse separately, because RCA had a consent decree involving patents in 1958 and thus might have changed its behavior. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Share of Communication Patents Measured with NBER Technology Subcategories

Figure A.6: Share of Communication Patents
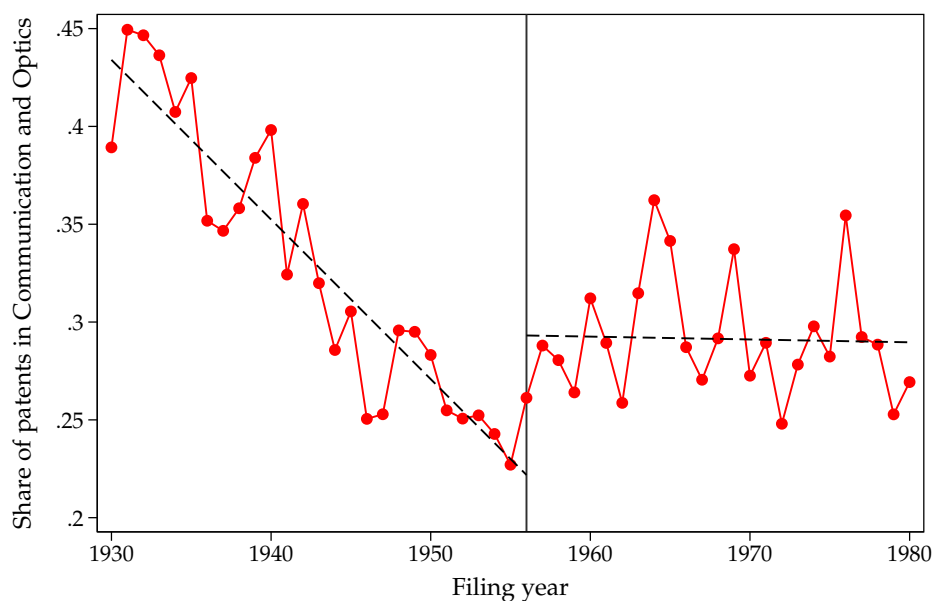
Notes: This figure shows the share of patents related to communication relative to all patents filed by Bell. We define technologies related to communication as the NBER subcategories "Communication" and "Optics" (Hall et al., 2001). We include "Optics" because after the invention of the laser at Bell Labs in 1958, Bell officials predicted correctly that optics might be crucial for the future of communication (Gertner, 2012, p. 253).

Effect for different definitions of small and young assignees

In Figure A.7 we estimate the main treatment coefficient separately for citations of different size and age groups of assignees. We find that the effect is driven mainly by companies and individual inventors without patents before 1949 and companies and individual inventors that are less than one year old at the time of the citations.

Figure A.7: Sample Split by Characteristics of Citing Firm

(a) By size of patent portfolio in 1949

(b) By age of company at citation
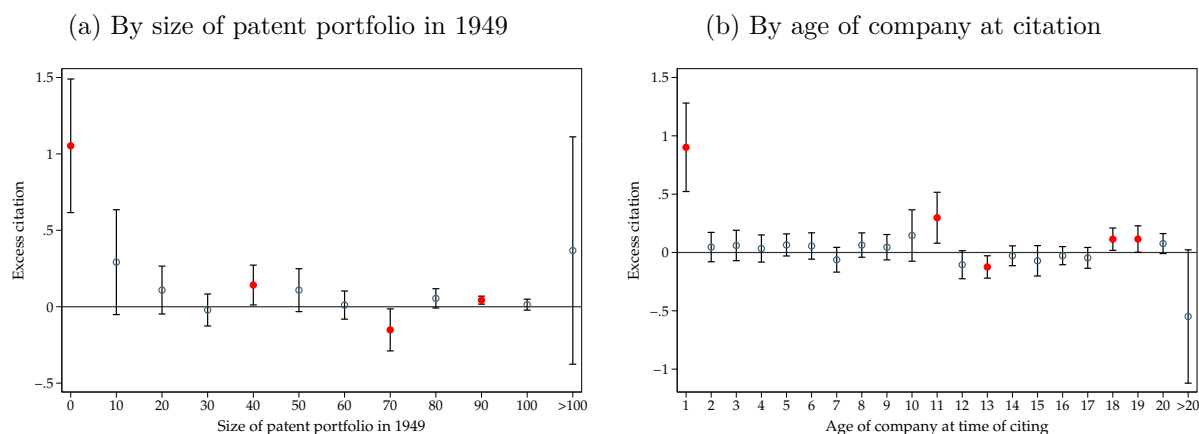
Notes: These Sub-figures show results from a difference-in-differences estimation with the years 1949-1954 as pre-treatment period and 1955-1960 as treatment period, controlling for year fixed effects. As dependent variable we use all citations by companies other than the filing companies with a specific size of their patent portfolio (Sub-Figure a) and a specific company age (b) as indicated in the figure. As control patents we use all patents that were published in the U.S. matched by publication year, primary USPC technology class, and the number of citations up to 1949. The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

A.2 Appendix to Section 1.5

Effect by NBER Technology Subcategory

In this section we estimate our main treatment effect separately for citations of patents in different NBER technology subcategories. The results are reported in Figure A.8. The increase in citations comes mainly from technologies related to electrical components, in particular in "Electrical Devices". Yet, there is no increase in citations by patents in the subcategory of "Communication". These results corroborate the finding in our main text that there is no increase in follow-on innovation in industries concerned with production of communication equipment, the core business of Bell.

No Lack of Follow-on Innovation in Telecommunications

This section presents evidence that the null effect in telecommunications was not due to a lack in potential follow-on innovation in the telecommunications market. To do this

Figure A.8: Effect of Compulsory Licensing on Subsequent Citations By NBER Technological Subcategory
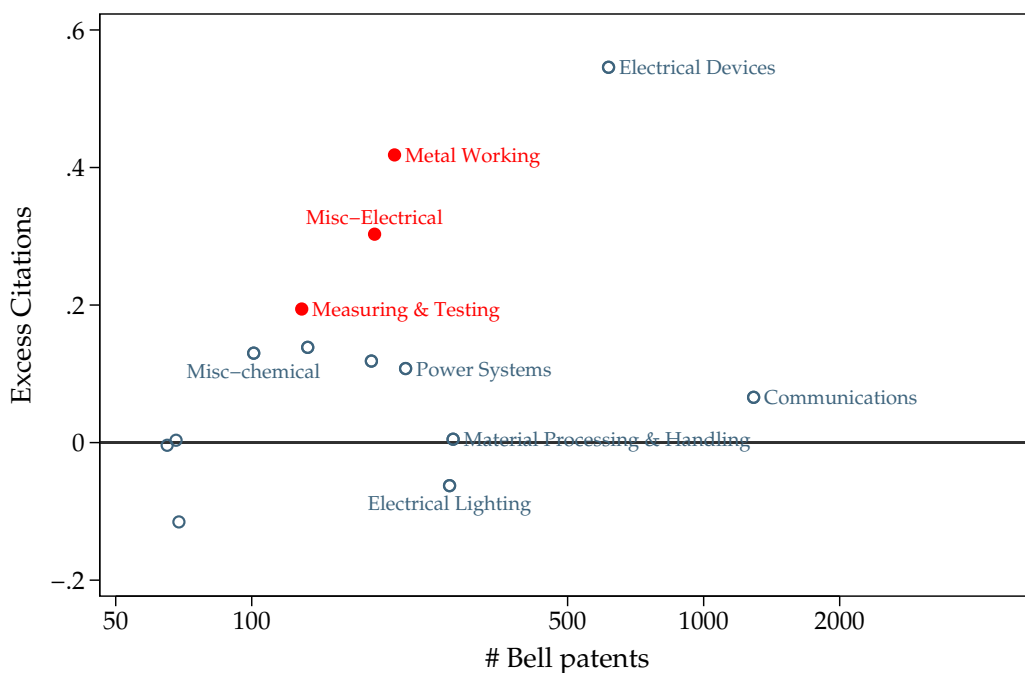
Notes: This figure shows difference-in-differences estimates of the impact of the consent decree on citations from patents in different NBER technological subcategories, controlling for year fixed effects. As dependent variable we use all citations by companies other than the filing company. As control patents we use all patents that were published in the U.S., matched by publication year, primary USPC technology class, and the number of citations up to 1949. A solid circle means that the coefficient is significant at the 10% level. We split the citing patents by NBER technology subcategory following Hall et al. (2001). The data are from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

Figure A.9: Number of Citations to Bell Patents Inside and Outside of Communication

(a) Up to 1949	(b) All Bell Patents
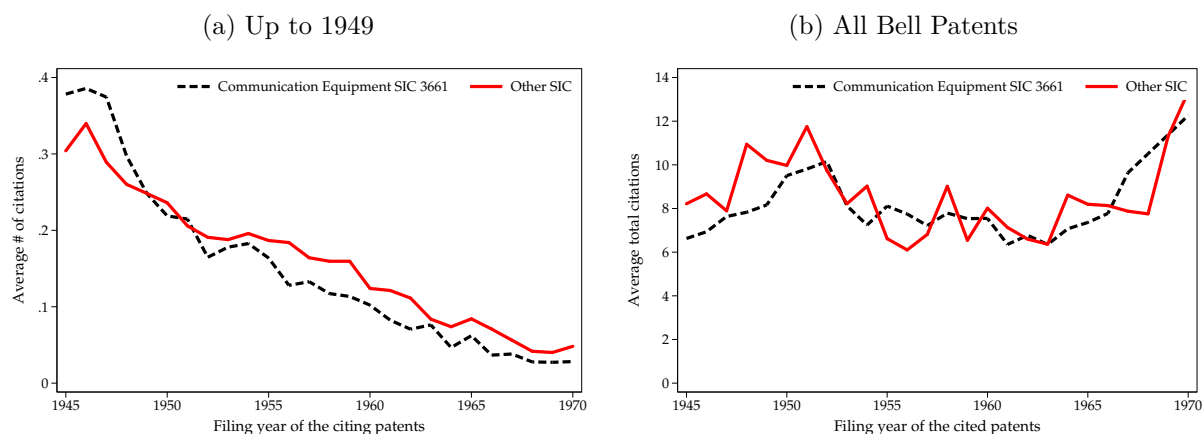
Notes: Sub-figure (a) shows the average number of citations per year for all Bell patents that are most likely used in the production of communication equipment (SIC 3661) and that are used in any other industry. To classify a patent by its most likely industry, we use the data of Kerr (2008) to assign to each USPC class the most likely four-digit SIC industry in which it is used. Sub-figure (b) shows the total number of citations to Bell patents inside and outside of telecommunication filed in a particular year. In this graph we use total citations, the sum of citations from other companies and from Bell to its own patents. The data stem from the Worldwide Patent Statistical Database (PATSTAT) of the European Patent Office.

we look at the total number of citations, the sum of citations of other companies and self-citations, to Bell patents inside and outside of telecommunications. In Subfigure (a) of Figure A.9 we plot the average number of total citations to Bell patents related to communication and related to other fields. We use the concordance of Kerr (2008) to assign to each Bell patent the most likely SIC code.

We find that the total number of citations to telecommunications patents of Bell were at least as high as to patents outside of communication. This speaks against a low quality of compulsorily licensed patents as a reason for the lack in follow-on innovation in telecommunications. In Subfigure (b) we show that the total number of patent citations to Bell's patents inside and outside of telecommunications were also almost identical before and after the consent decree. This suggests that after the consent decree the potential for follow-on innovation was not significantly lower in telecommunications than in other fields.

Appendix B

Appendix to Chapter 2

Table B.1: List of All Patent Libraries

City, State	Name of Library	Year
Albany, New York	New York State Library Cultural Education Center	1870
Boston, Massachusetts	Boston Public Library	1870
Columbus, Ohio	Science and Engineering Library. Ohio State University	1870
Los Angeles, California	Los Angeles Public Library	1870
New York, New York	New York Public Library	1870
St. Louis, Missouri	St. Louis Public Library	1870
Buffalo, New York	Buffalo and Erie County Public Library	1871
Cincinnati, Ohio	The Public Library of Cincinnati and Hamilton County	1871
Detroit, Michigan	Great Lakes Patent and Trademark Center. Detroit Public Library	1871
Chicago, Illinois	Chicago Public Library	1876
Newark, New Jersey	Newark Public Library	1880
Cleveland, Ohio	Cleveland Public Library	1890
Providence, Rhode Island	Providence Public Library	1901
Pittsburgh, Pennsylvania	The Carnegie Library of Pittsburgh	1902
Toledo, Ohio	Toledo/Lucas County Public Library	1934
Atlanta, Georgia	Library and Information Center. Georgia Institute of Technology	1946
Kansas City, Missouri	Linda Hall Library	1946
Milwaukee, Wisconsin	Milwaukee Public Library	1949
Stillwater, Oklahoma	Patent and Trademark Library. Oklahoma State University	1956
Sunnyvale, California	Sunnyvale Center for Innovation, Invention & Ideas (SC[I]3). Sunnyvale Public Library	1963
Madison, Wisconsin	Kurt F. Wendt Library. University of Wisconsin-Madison	1976
Birmingham, Alabama	Birmingham Public Library	1977
Dallas, Texas	Dallas Public Library	1977
Denver, Colorado	Denver Public Library	1977
Houston, Texas	Fondren Library. Rice University	1977
Raleigh, North Carolina	D.H. Hill Library. North Carolina State University	1977
Seattle, Washington	Engineering Library. University of Washington	1977
Lincoln, Nebraska	Engineering Library. University of Nebraska, Lincoln	1978
Sacramento, California	California State Library	1979
University Park, Pennsylvania	Schreyer Business Library. Paterno Library. Pennsylvania State Library	1979
Minneapolis, Minnesota	Minneapolis Public Library	1980
Newark, Delaware	University of Delaware Library	1980
Baton Rouge, Louisiana	Troy H. Middleton Library. Louisiana State University	1981
Albuquerque, New Mexico	Centennial Science and Engineering Library. The University of New Mexico	1983
Ann Arbor, Michigan	Media Union Library. The University of Michigan	1983
Auburn, Alabama	Ralph Brown Draughon Library. Auburn University	1983
Austin, Texas	McKinney Engineering Library. The University of Texas at Austin	1983
College Station, Texas	Sterling C. Evans Library. Texas A&M University	1983

Table B.2: List of All Patent Libraries (Continued)

City, State	Name of Library	Year
Indianapolis, Indiana	Indianapolis-Marion County Public Library	1983
Moscow, Idaho	University of Idaho Library	1983
Reno, Nevada	University Library. University of Nevada-Reno	1983
Amherst, Massachusetts	Physical Sciences and Engineering Library. University of Massachusetts	1984
Anchorage, Alaska	Z. J. Loussac Public Library. Anchorage Municipal Libraries	1984
Butte, Montana	Montana Tech Library of the University of Montana	1984
College Park, Maryland	Engineering and Physical Sciences Library. University of Maryland	1984
Fort Lauderdale, Florida	Broward County Main Library	1984
Miami, Florida	Miami-Dade Public Library System	1984
Salt Lake City, Utah	Marriott Library. University of Utah	1984
San Diego, California	San Diego Public Library	1984
Springfield, Illinois	Illinois State Library	1984
Little Rock, Arkansas	Arkansas State Library	1985
Nashville, Tennessee	Stevenson Science and Engineering Library. Vanderbilt	1985
Richmond, Virginia	James Branch Cabell Library. Virginia Commonwealth University	1985
Philadelphia, Pennsylvania	The Free Library of Philadelphia	1986
Washington, District of Columbia	Founders Library. Howard University	1986
Des Moines, Iowa	State Library of Iowa	1988
Louisville, Kentucky	Louisville Free Public Library	1988
Orlando, Florida	University of Central Florida Libraries	1988
Honolulu, Hawaii	Hawaii State Library	1989
Piscataway, New Jersey	Library of Science and Medicine. Rutgers University	1989
Grand Forks, North Dakota	Chester Fritz Library. University of North Dakota	1990
Jackson, Mississippi	Mississippi Library Commission	1990
Tampa, Florida	Patent Library. Tampa Campus Library. University of South Florida	1990
Wichita, Kansas	Ablah Library. Wichita State University	1991
Big Rapids, Michigan	Abigail S. Timme Library. Ferris State Library	1991
Morgantown, West Virginia	Evansdale Library. West Virginia University	1991
West Lafayette, Indiana	Siegesmund Engineering Library. Purdue University	1991
Clemson, South Carolina	R. M. Cooper Library. Clemson University	1992
Orono, Maine	Raymond H. Fogler Library. University of Maine	1993
Rapid City, South Dakota	Devereaux Library. South Dakota School of Mines and Technology	1994
San Francisco, California	San Francisco Public Library	1994
Akron, Ohio	Akron-Summit County Public Library	1995
Lubbock, Texas	Texas Tech University Library	1995
Mayaguez, Puerto Rico	General Library. University of Puerto Rico-Mayaguez	1995
Portland, Oregon	Paul L. Boley Law Library. Lewis & Clark Law School	1995
Burlington, Vermont	Bailey/Howe Library	1996
Concord, New Hampshire	New Hampshire State Library	1996
Hartford, Connecticut	Hartford Public Library	1997
New Haven,Connecticut	New Haven Free Public Library	1997
Stony Brook, New York	Engineering Library. Melville Library SUNY at Stony Brook	1997
Las Vegas, Nevada	Las Vegas Clark County Library District	1999
Rochester, New York	Central Library of Rochester and Monroe County	1999
Bayamon, Puerto Rico	Learning Resources Center. University of Puerto Rico-Bayamon Campus	2000
Dayton, Ohio	Paul Laurence Dunbar Library. Wright State University	2000
San Antonio, Texas	San Antonio Public Library	2000
Cheyenne, Wyoming	Wyoming State Library	2001

Table B.3: Libraries Not Used by Sample Restriction

Restriction	Libraries	
	not used	in sample
None	0	83
PTDL is not FDL	13	70
Medium and large library	3	67
Opening year in sample period (current patent data)	23	44
Control FDL within state	3	41
Rochester, Burlington, Puerto Rico	3	38

Appendix C

Appendix to Chapter 3

C.1 Matching Researchers to their Scientific Output and Identifying the Treatment Group

The Scopus database groups publications by authors, which means that it is sufficient to match the researchers from the appointment lists to their respective counterpart in Scopus.[1] After that, collecting the researchers' publications and the corresponding forward citations happens within Scopus. We apply a multi-layer matching procedure by first using researcher-specific information from the appointment lists, and second publication-specific information from Scopus get more precise matches.

In the following, the information used in the first step of the matching procedure are listed. A Scopus person is only considered a match if all the following variables concur.

Name: Full first name and full last name of the researchers from the appointment lists are used, if available. For 3.6% of the researchers we only know the last name, and for 5% we only know the initial of the first name. These missing information are largely due to the early lists in the 1950s and 60s.

Place of origin: For 98.4% of the researchers we have information about the place of employment at the time of application. We match these places to the standardized city

---

[1]This section stems from the companion paper, Watzinger et al. (2017b).

and institution names in Scopus.

Field of research: From the appointment lists, we have information about the faculty and the exact chair a researcher applied to. We match these data to the field of research classification from Scopus.

Years of publications: After exhausting the researcher-specific information, we extract the corresponding publications from Scopus. This gives us additional information, specifically the time of activity of a researcher. We only consider matches where the year of the appointment list falls within the time of activity, which is defined by the earliest and latest publication of a Scopus researcher.

Manual search: For the remaining unmatched researchers, and for matches that seem implausible (e.g. multiple matches), we try to identify Scopus matches by manually searching the database for i.e. known publications.

To identify the treatment group, i.e., the researchers of the appointment lists who actually accepted the position and moved to the new university, we mainly use two approaches: First, we compare the names of the researchers to the names found in the course catalog of the university in question. Second, we use the affiliation information from the Scopus publications to identify the one researcher with the new university as affiliation in the years after the move. For the remaining lists where no treated researcher could be identified and for identified treated researchers that seem implausible (e.g., multiple treated researchers within one list), manual searches are conducted.

C.2   Additional Results

Figure C.1: Intent-to-Treat Effect of Moving on Scientist Productivity
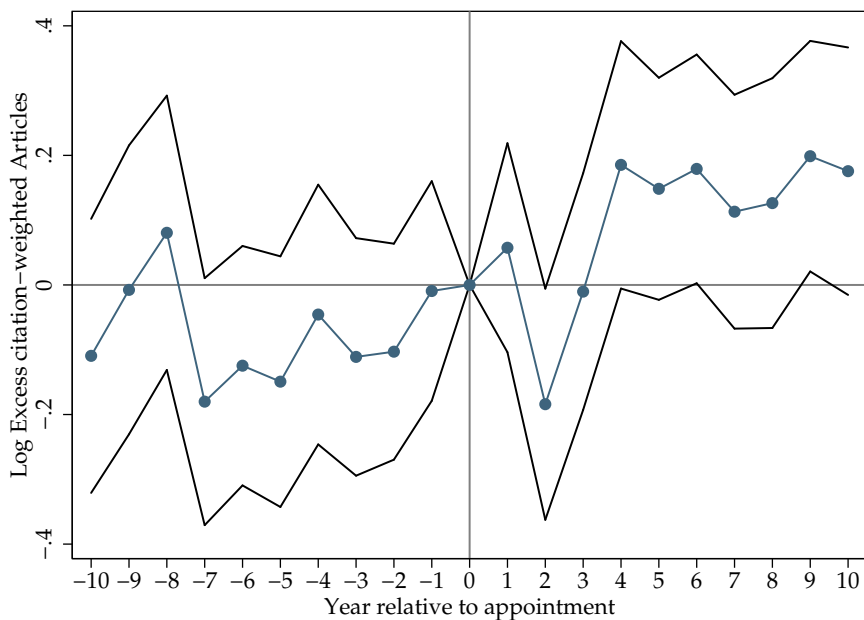
Notes: This graph shows the estimated number of yearly log excess publications by highest-ranked researchers, relative to the year of the move. Publications are weighted by the number of authors and by the number of follow-on citations. To arrive at these estimates, we regress the yearly log number of weighted publications by researchers on the appointment lists on an indicator variable equal to one if the researcher is the ranked first on the appointment list (interacted with time dummies), year fixed effects, and appointment list fixed effects. The dark blue line represents the 90% confidence bands for the estimated coefficients. Standard errors allow for clustering at the appointment list level. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

Table C.1: Alternative Estimation Methods

Dep. Var.:	Publications					
	Cit.- & Auth.-Wght.			Auth.-Wght.	Raw	
	(1)	(2)	(3)	(4)	(5)	(6)
Mover	0.07	2.32	0.00	0.07	0.03	0.03
	(0.07)	(2.88)	(0.08)	(0.05)	(0.06)	(0.06)
Post	-0.03	-0.55	-0.04	0.06*	0.18***	0.10***
	(0.04)	(1.49)	(0.05)	(0.03)	(0.03)	(0.03)
Post x Move	0.19***	5.50**	0.18**	0.13***	0.20***	0.13***
	(0.06)	(2.38)	(0.07)	(0.05)	(0.06)	(0.04)
Constant	1.87***	28.47***				0.41***
	(0.03)	(1.14)				(0.08)
List FE	Yes	Yes	Yes	Yes	Yes	Yes
Estimation	Baseline	Baseline	Poisson	Poisson	Poisson	Neg. Bin.
Dep Var	Ln+1	Levels	Levels	Levels	Levels	Levels
Adj R2	0.45	0.23				
Clust. (Lists)	317	317				312
Log-Likelihood	-16854.86	-55694.16	-2.49e+05	-11858.91	-24669.05	-19959.10
Observations	10000	10000	9880	9900	9900	9900

Notes: This table shows regressions using alternative estimation methods on the author-by-year level. The data comprises the 5 years before and 5 years after the appointment. The dependent variable is the yearly number author-and citation-weighted publications of authors on the appointment list in logs in column (1) and in levels in columns (2), (3), and (6). It is the author-weighted number of publications in column (4) and the raw number of publications in column (5). Columns (1) and (2) use a linear difference-in-differences model as in the baseline specifications. Standard errors allowing for clustering on the appointment list level are in parentheses. Columns (3) through (5) use fixed-effects poisson regression with robust standard errors (see, e.g., Hausman et al., 1984; Wooldridge, 1999). These standard errors allow for clustering at the appointment list level. Column (6) uses fixed-effects negative binomial regression. Standard errors are bootstrapped with 1'000 replications. In both poisson and negative binomial regressions, observations with zero outcome are dropped. ***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The data stems from the universe of appointment lists of one large university in Germany between 1950 and 2005. Publication data stems from Scopus.

Appendix D

Appendix to Chapter 4

Table D.1: Recession at Career Start and Teacher Math Effectiveness (Quantile Regressions)

	Dependent variable: VAM in math						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Recession	0.029	0.096***	0.102***	0.088***	0.116***	0.139***	0.175***
	(0.035)	(0.027)	(0.031)	(0.025)	(0.026)	(0.024)	(0.037)
Quantile	5	10	25	50	75	90	95
Obs. (Teachers)	32941	32941	32941	32941	32941	32941	32941
R^2	0.020	0.021	0.021	0.021	0.021	0.015	0.008

Notes: Coefficients from separate quantile regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on NBER recession indicator at career start at different quantiles of the VAM distribution. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** $p < 1\%$, ** $p < 5\%$, * $p < 10\%$

Table D.2: Recession at Career Start and Teacher Math Effectiveness (Alternative VAMs)

	Dependent variable: Various VAMs in math				
	(1)	(2)	(3)	(4)	(5)
Recession	0.110***	0.090***	0.059***	0.092***	0.083***
	(0.023)	(0.022)	(0.017)	(0.029)	(0.027)
Fixed effects (in VAM model)	none	school	school-year	none	none
Weights	none	none	none	student obs.	teacher obs.
Clusters (career start years)	60	60	60	60	60
Obs. (teachers)	32941	32941	32941	32941	32941
R^2	0.022	0.018	0.014	0.019	0.020

Notes: Coefficients from separate regressions of different VAMs in math (controlling for yearly experience dummies up to 30 years) on NBER recession indicator at career start. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

Table D.3: Recession at Career Start and Teacher Math Effectiveness (Further Business Cycle Measures)

	Dependent variable: VAM in math					
	(1)	(2)	(3)	(4)	(5)	(6)
Unemp. (College)	0.052**					
	(0.022)					
Unemp. diff. (College)		0.083***				
		(0.015)				
Unemp. (Nat.)			0.031***			
			(0.010)			
Unemp. diff. (Nat.)				0.046***		
				(0.012)		
Unempl. (FL)					0.025***	
					(0.008)	
Unempl. diff. (FL)						0.024**
						(0.011)
Clusters (Career start years)	40	39	60	60	53	52
Obs. (Teachers)	32402	32244	32941	32941	32928	32923
R^2	0.021	0.022	0.021	0.022	0.021	0.021

Notes: Coefficients from separate regressions of VAM in math (controlling for yearly experience dummies up to 30 years) on alternative business cycle measures at career start. Unemployment (college) refers to BLS unemployment rates of college graduates (4 years and above until 1991, degree holders after 1991) and are available after 1969. Other unemployment rates are not graduate-specific. All unemployment rates are from the BLS. Teacher and student data stems from the Florida Department of Education. Standard errors in parentheses adjusted for clustering at the career start year level. Significance levels: *** p< 1%, ** p< 5%, * p< 10%

Bibliography

ACEMOGLU, D. AND U. AKCIGIT (2012): "Intellectual Property Rights Policy, Competition and Innovation," *Journal of the European Economic Association*, 10, 1–42.

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4B, 1043–1171.

AGHION, P., C. HARRIS, P. HOWITT, AND J. VICKERS (2001): "Competition, Imitation and Growth with Step-by-step Innovation," *Review of Economic Studies*, 68, 467–492.

AGRAWAL, A., J. MCHALE, AND A. OETTL (2017): "How Stars Matter: Recruiting and Peer Effects in Evolutionary Biology," *Research Policy*, 46, 853–867.

AKCIGIT, U. AND W. R. KERR (forthcoming): "Growth Through Heterogeneous Innovations," *Journal of Political Economy*.

ALCACER, J. AND M. GITTELMAN (2006): "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations," *Review of Economics and Statistics*, 88, 774–779.

ALCACER, J., M. GITTELMAN, AND B. SAMPAT (2009): "Applicant and Examiner Citations in US Patents: An Overview and Analysis," *Research Policy*, 38, 415–427.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

ANTITRUST SUBCOMMITTEE (1958): "Consent Decree Program of the Department of Justice. American Telephone & Tel. Subcommittee - Hearings before the Antitrust

Subcommittee," *U.S. House Committee on the Judiciary, Committee on the Judiciary Serial No. 9.*

——— (1959): "Consent Decree Program of the Department of Justice," *Committee on the Judiciary, House of Representatives, Eighty-Sixth Congress, First Session.*

ARORA, A., S. BELENZON, AND A. PATACCONI (forthcoming): "The Decline of Science in Corporate R&D," *Strategic Management Journal.*

ARORA, A., M. CECCAGNOLI, AND W. M. COHEN (2008): "R&D and the Patent Premium," *International Journal of Industrial Organization*, 26, 1153–1179.

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

ASHRAF, N., O. BANDIERA, AND S. S. LEE (2016): "Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery," *Mimeo.*

AUDRETSCH, D. B. AND M. P. FELDMAN (2004): "Knowledge Spillovers and the Geography of Innovation," in *Handbook of Regional and Urban Economics*, ed. by G. Duranton, J. V. Henderson, and W. C. Strange, Elsevier, vol. 4, 2713–2739.

AUTOR, D. H. (2014): "Skills, Education, and the Rise of Earnings Inequality Among the "Other 99 Percent"," *Science*, 344, 843–851.

——— (2015): "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, 29, 3–30.

AZOULAY, P., I. GANGULI, AND J. G. ZIVIN (2017): "The Mobility of Elite Life Scientists: Professional and Personal Determinants," *Research Policy*, 46, 573–590.

AZOULAY, P., J. S. GRAFF ZIVIN, AND J. WANG (2010): "Superstar Extinction," *Quarterly Journal of Economics*, 125, 549–589.

BACHER-HICKS, A., T. J. KANE, AND D. O. STAIGER (2014): "Validating Teacher Effects Estimates Using Changes in Teacher Assignments in Los Angeles," *NBER Working Paper No. 20657.*

BIBLIOGRAPHY

BACOLOD, M. P. (2007): "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality," *Review of Economics and Statistics*, 89, 737–751.

BAEKER, A. (2015): "Non-tenured Post-doctoral Researchers' Job Mobility and Research Output: An Analysis of the Role of Research Discipline, Department Size, and Coauthors," *Research Policy*, 44, 634–650.

BAKER, J. B. (2012): "Exclusion as a Core Competition Concern," *Antitrust Law Journal*, 78, 527–589.

BEAUDRY, P., D. A. GREEN, AND B. M. SAND (2016): "The Great Reversal in the Demand for Skill and Cognitive Tasks," *Journal of Labor Economics*, 34, S199–S247.

BERMAN, J. AND J. PFLEEGER (1997): "Which Industries are Sensitive to Business Cycles?" *Monthly Labor Review*, 120, 19–25.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How Much Should We Trust Differences-in-differences Estimates?" *Quarterly Journal of Economics*, 119, 249–275.

BESLEY, T. AND M. GHATAK (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95, 616–636.

BIASI, B. AND P. MOSER (2016): "Effects of Copyrights on Science: Evidence from the WWII Book Republication Program," *Mimeo*.

BOEHM, M. AND M. WATZINGER (2015): "The Allocation of Talent over the Business Cycle and its Long-term Effect on Sectoral Productivity," *Economica*, 82, 892–911.

BORJAS, G. J. (1987): "Self-selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

——— (2002): "The Wage Structure and the Sorting of Workers into the Public Sector," *NBER Working Paper No. 9313*.

171

Bibliography

Boudreau, K., T. Brady, I. Ganguli, P. Gaule, E. Guinan, T. Hollenberg, and K. Lakhani (forthcoming): "A Field Experiment on Search Costs and the Formation of Scientific Collaborations," *Review of Economics and Statistics.*

Britton, J. and C. Propper (2016): "Teacher Pay and School Productivity: Exploiting Wage Regulation," *Journal of Public Economics*, 133, 75–89.

Catalini, C. (forthcoming): "Microgeography and the Direction of Inventive Activity," *Management Science.*

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-added Estimates," *American Economic Review*, 104, 2593–2632.

——— (2014b): "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood," *American Economic Review*, 104, 2633–2679.

Chingos, M. M. and P. E. Peterson (2011): "It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness," *Economics of Education Review*, 30, 449–465.

Chingos, M. M. and M. R. West (2012): "Do More Effective Teachers Earn More Outside the Classroom?" *Education Finance and Policy*, 7, 8–43.

Cohen, W. M., A. Goto, A. Nagata, R. R. Nelson, and J. P. Walsh (2002): "R&D Spillovers, Patents and the Incentives to Innovate in Japan and the United States," *Research Policy*, 31, 1349–1367.

Corcoran, S. P., W. N. Evans, and R. M. Schwab (2004): "Changing Labor-market Opportunities for Women and the Quality of Teachers, 1957-2000," *American Economic Review, Papers and Proceedings*, 94, 230–235.

Cowan, J. C., D. Goldhaber, K. Hayes, and R. Theobald (2016): "Missing Elements in the Discussion of Teacher Shortages," *CALDER Explainer, National Center for the Analysis of Longitudinal Data in Education Research.*

DE REE, J., K. MURALIDHARAN, M. PRADHAN, AND H. ROGERS (forthcoming): "Double for Nothing? The Effect of Unconditional Teachers' Salary Increases on Performance," *Quarterly Journal of Economics*.

DELRAHIM, M. (2004): "Forcing Firms to Share the Sandbox: Compulsory Licensing of Intellectual Property Rights and Antitrust," *European Business Law Review*, 15, 1059–1069.

DEMING, D. J. (forthcoming): "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*.

DOLTON, P. J. (1990): "The Economics of UK Teacher Supply: The Graduate's Decision," *Economic Journal*, 100, 91–104.

——— (2006): "Teacher Supply," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek and F. Welch, Elsevier, vol. 2, chap. 19, 1079–1161.

DOLTON, P. J. AND O. D. MARCENARO-GUTIERREZ (2011): "If you Pay Peanuts do you get Monkeys? A Cross-Country Analysis of Teacher Pay and Pupil Performance," *Economic Policy*, 26, 5–55.

DRUCKER, P. F. (1999): "Knowledge-worker Productivity: The Biggest Challenge," *California Management Review*, 41, 79–94.

FALCH, T., K. JOHANSEN, AND B. STROM (2009): "Teacher Shortages and the Business Cycle," *Labour Economics*, 16, 648–658.

FIGLIO, D. (1997): "Teacher Salaries and Teacher Quality," *Economics Letters*, 55, 267–271.

FITZPATRICK, M. D. (2015): "How Much Do Public School Teachers Value Their Pension Benefits?" *American Economic Journal: Economic Policy*, 7, 165–188.

FLORIDA DEPARTMENT OF EDUCATION (1986): "Teacher Supply and Demand in Florida: Fifth Annual Report." *Strategy Planning and Management Information Systems Section, Tallahassee, FL.*

BIBLIOGRAPHY

FOSTER, L., J. HALTIWANGER, AND C. J. KRIZAN (2006): "Market Selection, Reallocation, and Restructuring in the US Retail Trade Sector in the 1990s," *Review of Economics and Statistics*, 88, 748–758.

FRANZONI, C., G. SCELLATO, AND P. STEPHAN (2014): "The Mover's Advantage: The Superior Performance of Migrant Scientists," *Economics Letters*, 122, 89–93.

FURMAN, J. L., M. NAGLER, AND M. WATZINGER (2017): "Disclosure and Cumulative Innovation: Evidence from the Patent Depository Library Program," *Mimeo*.

FURMAN, J. L. AND S. STERN (2011): "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research," *American Economic Review*, 101, 1933–1963.

GALASSO, A. (2012): "Broad Cross-license Negotiations," *Journal of Economics & Management Strategy*, 21, 873–911.

GALASSO, A. AND M. SCHANKERMAN (2015): "Patents and Cumulative Innovation: Causal Evidence from the Courts," *Quarterly Journal of Economics*, 130, 317–369.

——— (forthcoming): "Patent Rights, Innovation, and Firm Exit," *RAND Journal of Economics*.

GAMBARDELLA, A., D. HARHOFF, AND S. NAGAOKA (2011): "The Social Value of Patent Disclosure," *Mimeo*.

GANGULI, I. (2015a): "Emigrate or Collaborate? Evidence on Location and Scientific Productivity from the Soviet 'Brain Drain'," *Mimeo*.

——— (2015b): "Immigration and Ideas: What Did Russian Scientists Bring to the United States?" *Journal of Labor Economics*, 33, S257–S288.

GANS, J. S., D. H. HSU, AND S. STERN (2002): "When Does Start-up Innovation Spur the Gale of Creative Destruction?" *RAND Journal of Economics*, 33, 571–586.

——— (2008): "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence From Patent Grant Delays," *Management Science*, 54, 982–997.

GANS, J. S. AND S. STERN (2003): "The Product Market and the Market for Ideas: Commercialization Strategies for Technology Entrepreneurs," *Research Policy*, 32, 333–350.

GERTNER, J. (2012): *The Idea Factory: Bell Labs and the Great Age of American Innovation*, Penguin.

GIULIANO, P. AND A. SPILIMBERGO (2014): "Growing Up in a Recession," *Review of Economic Studies*, 81, 787–817.

GLAESER, E. L., H. D. KALLAL, J. A. SCHEINKMAN, AND A. SHLEIFER (1992): "Growth in Cities," *Journal of Political Economy*, 100, 1126–1152.

GOLDHABER, D. AND J. WALCH (2013): "Rhetoric Versus Reality: Is the Academic Caliber of the Teacher Workforce Changing?" *CEDR Working Paper 2013-4*.

GRAHAM, S. AND D. HEGDE (2015): "Disclosing Patents' Secrets," *Science*, 347, 236–237.

GRINDLEY, P. C. AND D. J. TEECE (1997): "Licensing and Cross-Licensing in Semiconductors and Electronics," *California Management Review*, 39, 8–41.

HALL, B. H. AND D. HARHOFF (2012): "Recent Research on the Economics of Patents," *Annual Review of Economics*, 4, 541–565.

HALL, B. H., A. B. JAFFE, AND M. TRAJTENBERG (2001): "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," *NBER Working Paper No. 8498*.

HANUSHEK, E. A. (2011): "The Economic Value of Higher Teacher Quality," *Economics of Education Review*, 30, 466–479.

HANUSHEK, E. A. AND R. R. PACE (1995): "Who Chooses To Teach (and Why)?" *Economics of Education Review*, 14, 101–117.

HANUSHEK, E. A., M. PIOPIUNIK, AND S. WIEDERHOLD (2014): "The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance," *NBER Working Paper No. 20727*.

Hanushek, E. A. and S. G. Rivkin (2012): "The Distribution of Teacher Quality and Implications for Policy," *Annual Review of Economics*, 4, 131–157.

Hanushek, E. A., J. Ruhose, and L. Woessman (forthcoming): "Knowledge Capital and Aggregate Income Differences: Development Accounting for US States," *American Economic Journal: Macroeconomics*.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015): "Returns to Skills Around the World: Evidence from PIAAC," *European Economic Review*, 73, 103–130.

Hanushek, E. A. and L. Woessmann (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–668.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52, 909–938.

Heckman, J., N. Hohmann, J. Smith, and M. Khoo (2000): "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment," *Quarterly Journal of Economics*, 115, 651–694.

Hegde, D. and H. Luo (2017): "Patent Publication and the Market for Ideas," *Management Science*.

Helpman, E. (1998): *General Purpose Technologies and Economic Growth*, MIT Press.

Hoisl, K. (2007): "Tracing Mobile Inventors - The Causality Between Inventor Mobility and Inventor Productivity," *Research Policy*, 36, 619–636.

——— (2009): "Does Mobility Increase the Productivity of Inventors? New Evidence from a Quasi-Experimental Design," *Journal of Technology Transfer*, 34, 212–225.

Holbrook, D., W. M. Cohen, D. A. Hounshell, and S. Klepper (2000): "The Nature, Sources, and Consequences of Firm Differences in the Early History of the Semiconductor Industry," *Strategic Management Journal*, 21, 1017–1041.

HOXBY, C. M. AND A. LEIGH (2004): "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States," *American Economic Review, Papers and Proceedings*, 94, 236–240.

HOYNES, H., D. L. MILLER, AND J. SCHALLER (2012): "Who Suffers During Recessions?" *Journal of Economic Perspectives*, 26, 27–48.

IACUS, S. M., G. KING, AND G. PORRO (2009): "CEM: Software for Coarsened Exact Matching," *Journal of Statistical Software*, 30, 1–27.

———— (2012): "Causal Inference Without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 20, 1–24.

IMBENS, G. W. AND D. B. RUBIN (2015): *Causal Inference in Statistics, Social, and Biomedical Science: An Introduction*, Cambridge University Press.

JACKSON, C. K. (2012): "Non-cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina," *NBER Working Paper No. 18624*.

JACKSON, C. K. AND E. BRUEGMANN (2009): "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers," *American Economic Journal: Applied Economics*, 1, 85–108.

JACKSON, C. K., J. E. ROCKOFF, AND D. O. STAIGER (2014): "Teacher Effects and Teacher-related Policies," *Annual Review of Economics*, 6, 801–825.

JAFFE, A. B. AND J. LERNER (2011): *Innovation and its Discontents: How our Broken Patent System is Endangering Innovation and Progress, and What to do About it*, Princeton University Press.

JENDA, C. A. (2005): "Patent and Trademark Depository Libraries and the United States Patent and Trademark Office: A Model for Information Dissemination," *Resource Sharing & Information Networks*, 18, 183–201.

JONES, B. F. (2009): "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *Review of Economic Studies*, 76, 283–317.

——— (2010): "Age and Great Invention," *Review of Economics and Statistics*, 92, 1–14.

JONES, B. F. AND B. A. WEINBERG (2011): "Age Dynamics in Scientific Creativity," *Proceedings of the National Academy of Sciences*, 108, 18910–18914.

JONES, C. I. (1995): "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103, 759–784.

——— (2016): "The Facts of Economic Growth," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 3–69.

JONES, C. I. AND P. M. ROMER (2010): "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, 2, 224–245.

JONES, J. AND J. SCHMITT (2014): "A College Degree is No Guarantee," *Working Paper, Center for Economic and Policy Research*.

KAHN, L. B. (2010): "The Long-term Labor Market Consequences of Graduating from College in a Bad Economy," *Labour Economics*, 17, 303–316.

KANE, T. J., D. F. MCCAFFREY, T. MILLER, AND D. O. STAIGER (2013): "Have We Identified Effective Teachers?" *MET Project Research Paper, Bill & Melinda Gates Foundation.*

KANE, T. J. AND D. O. STAIGER (2008): "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," *NBER Working Paper No. 14607.*

KERR, W. R. (2008): "Ethnic Scientific Communities and International Technology Diffusion," *Review of Economics and Statistics*, 90, 518–537.

KOGAN, L., D. PAPANIKOLAOU, A. SERU, AND N. STOFFMAN (2017): "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, 132, 665–712.

KOPELMAN, J. L. AND H. S. ROSEN (2016): "Are Public Sector Jobs Recession-proof? Were They Ever?" *Public Finance Review*, 44, 370–396.

BIBLIOGRAPHY

KORTUM, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65, 1389–1419.

KRUEGER, A. B. (1988): "The Determinants of Queues for Federal Jobs," *Industrial and Labor Relations Review*, 41, 567–581.

LAKDAWALLA, D. (2006): "The Economics of Teacher Quality," *Journal of Law and Economics*, 49, 285–329.

LANJOUW, J. O. AND M. SCHANKERMAN (2004): "Protecting Intellectual Property Rights: Are Small Firms Handicapped?" *Journal of Law and Economics*, 47, 45–74.

LERNER, J. (1994): "The Importance of Patent Scope: An Empirical Analysis," *RAND Journal of Economics*, 25, 319–333.

LI, G.-C., R. LAI, A. D'AMOUR, D. M. DOOLIN, Y. SUN, V. I. TORVIK, Z. Y. AMY, AND L. FLEMING (2014): "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975-2010)," *Research Policy*, 43, 941–955.

LOEB, S. AND M. E. PAGE (2000): "Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-pecuniary Variation," *Review of Economics and Statistics*, 82, 393–408.

MACHLUP, F. (1958): *An Economic Review of the Patent System*, 15, US Government Printing Office.

MANKIW, N. G., D. ROMER, AND D. N. WEIL. (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407–437.

MANSO, G. (2011): "Motivating Innovation," *Journal of Finance*, 66, 1823–1860.

MCKENZIE, D., S. STILLMAN, AND J. GIBSON (2010): "How Important is Selection? Experimental vs. Non-experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, 8, 913–945.

MOE, T. M. (2006): "Quality Teachers," in *Reforming Education in Florida*, ed. by P. E. Peterson, Hoover Institution Press, 135–148.

MOKYR, J. (2002): *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton University Press.

MOSER, P. AND A. VOENA (2012): "Compulsory Licensing: Evidence from the Trading with the Enemy Act," *American Economic Review*, 102, 396–427.

MOSER, P., A. VOENA, AND F. WALDINGER (2014): "German Jewish Emigres and US Invention," *American Economic Review*, 104, 3222–3255.

MOWERY, D. C. (2011): "Federal Policy and the Development of Semiconductors, Computer Hardware, and Computer Software: A Policy Model for Climate Change R&D?" in *Accelerating Energy Innovation: Insights from Multiple Sectors*, ed. by R. M. Henderson and R. G. Newell, University of Chicago Press, 159–188.

MURNANE, R. J. AND B. R. PHILLIPS (1981): "Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance," *Economics of Education Review*, 1, 453–465.

MURPHY, K. M., A. SHLEIFER, AND R. W. VISHNY (1991): "The Allocation of Talent: Implications for Growth," *Quarterly Journal of Economics*, 106, 503–530.

MURRAY, F. AND S. STERN (2007): "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-commons Hypothesis," *Journal of Economic Behavior & Organization*, 63, 648–687.

NAGLER, M., M. PIOPIUNIK, AND M. R. WEST (2015): "Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness," *NBER Working Paper No. 21393*.

NATIONAL PUBLIC RADIO (2015): "Where Have All The Teachers Gone?" *http://www.npr.org/blogs/ed/2015/03/03/389282733/where-have-all-the-teachers-gone*, March 03.

NATIONAL SCIENCE BOARD (2016): *Science and Engineering Indicators 2016*, National Science Foundation (NSB-2016-1).

Neef, D. (1998): "Rethinking Economics in the Knowledge-based Economy," in *The Economic Impact of Knowledge*, ed. by D. Neef, G. A. Siesfeld, and J. Cefola, Butterworth Press, 3–16.

Nelson, R. R. (1959): "The Simple Economics of Basic Scientific Research," *Journal of Political Economy*, 67, 297–306.

——— (1962): "The Link Between Science and Invention: The Case of the Transistor," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton University Press, 549–584.

New York Times (2010): "Teachers Facing Weakest Market in Years," May 19.

——— (2015): "Fewer Top Graduates Want to Join Teach for America," February 6.

Noll, A. (1987): "Bell System R&D Activities," *Telecommunications Policy*, 11, 161 – 178.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012): "The Short- and Long-term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4, 1–29.

Oyer, P. (2006): "Initial Labor Market Conditions and Long-term Outcomes for Economists," *Journal of Economic Perspectives*, 20, 143–160.

——— (2008): "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *Journal of Finance*, 63, 2601–2628.

Papay, J. P. and M. A. Kraft (2015): "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement," *Journal of Public Economics*, 130, 105–119.

Powell, W. W. and K. Snellman (2004): "The Knowledge Economy," *Annual Review of Sociology*, 30, 199–220.

Rockoff, J. E. (2004): "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review, Papers and Proceedings*, 94, 247–252.

Bibliography

ROIN, B. (2005): "The Disclosure Function of the Patent System (or Lack Thereof)," *Harvard Law Review*, 118, 2007–2028.

ROMER, P. M. (1986): "Increasing Returns and Long-run Growth," *Journal of Political Economy*, 94, 1002–1037.

——— (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

——— (1994): "The Origins of Endogenous Growth," *Journal of Economic Perspectives*, 8, 3–22.

ROSENBERG, N. (1990): "Why do Firms do Basic Research (with their own Money)?" *Research Policy*, 19, 165–174.

ROTHSTEIN, J. (2017): "Measuring the Impact of Teachers: Comment," *American Economic Review*, 107, 1656–1684.

ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

SAMPAT, B. AND H. L. WILLIAMS (2015): "How do Patents Affect Follow-on Innovation? Evidence from the Human Genome," *NBER Working Paper No. 21666*.

SANTOS SILVA, J. M. C. AND S. TENREYRO (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88, 641–658.

SCOTCHMER, S. (1991): "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5, 29–41.

SCOTCHMER, S. AND J. GREEN (1990): "Novelty and Disclosure in Patent Law," *RAND Journal of Economics*, 21, 131–146.

SEGAL, I. AND M. D. WHINSTON (2007): "Antitrust in Innovative Industries," *American Economic Review*, 97, 1703–1730.

SHU, P. (2012): "The Long-term Impact of Business Cycles on Innovation: Evidence from the Massachusetts Institute of Technology," *Mimeo*.

SIMPKINS, J., M. ROZA, AND S. SIMBURG (2012): "What Happens to Teacher Salaries During a Recession?" *Center on Reinventing Public Education.*

SNEED, M. C. (1998): "125 years of patent information to the people: the US Patent and Trademark Depository Library Program, 22 May 1997," .

——— (2000): "Fully Disclosed yet Merely Descriptive: Intricacies of Training the Patent and Trademark Information Professional," *Journal of Library Administration*, 29, 59–78.

STEPHAN, P. E. (1996): "The Economics of Science," *Journal of Economic Literature*, 34, 1199–1235.

——— (2012): *How Economics Shapes Science*, Harvard University Press.

STERN, S. (2004): "Do Scientists Pay to Be Scientists?" *Management Science*, 50, 835–853.

STURIALE, J. E. (2011): "Compulsory Licensing of Intellectual Property as Merger Remedy: A Decision-theoretic Approach," *Louisiana Law Review*, 72, 605–646.

TEMIN, P. AND L. GALAMBOS (1987): *The Fall of the Bell System: A Study in Prices and Politics*, Cambridge University Press.

THE ECONOMIST (2000): "Who Owns the Knowledge Economy?" April 6.

——— (2005): "Brain Teasing," October 13.

TODD, P. E. AND K. I. WOLPIN (2003): "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, 113, F3–F33.

TRAJTENBERG, M. (2005): "Recombinant Ideas: The Mobility of Inventors and the Productivity of Research," *Presentation at the CEPR Conference, Munich, May 26-28, 2005.*

U.S. DEPARTMENT OF EDUCATION (2013): "Preparing and Credentialing the Nation's Teachers: The Secretary's Ninth Report on Teacher Quality." *Office of Postsecondary Education, Washington, DC.*

Bibliography

UZZI, B., S. MUKHERJEE, M. STRINGER, AND B. JONES (2013): "Atypical Combinations and Scientific Impact," *Science*, 342, 468–472.

WALL STREET JOURNAL (2016): "The Rise of Knowledge Workers Is Accelerating Despite the Threat of Automation," May 4.

WALLER, S. W. AND M. SAG (2014): "Promoting Innovation," *Iowa Law Review*, 100, 2223–2247.

WANG, J., R. VEUGELERS, AND P. STEPHAN (2017): "Bias Against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators," *Research Policy*, 46, 1416–1436.

WATZINGER, M., T. A. FACKLER, M. NAGLER, AND M. SCHNITZER (2017a): "How Antitrust Enforcement Can Spur Innovation: Bell Labs and the 1956 Consent Decree," *CEPR Discussion Paper No. 11793*.

WATZINGER, M., M. NAGLER, L. TREBER, AND M. SCHNITZER (2017b): "Mobility of Scientists and the Spread of Ideas," *Mimeo*.

WEITZMAN, M. L. (1998): "Recombinant Growth," *Quarterly Journal of Economics*, 113, 331–360.

WESSNER, C. W. ET AL. (2001): *Capitalizing on New Needs and New Opportunities: Government-industry Partnerships in Biotechnology and Information Technologies*, National Academies Press, chapter 2 available at http://www.nap.edu/read/10281/chapter/11.

WILLIAMS, H. L. (2013): "Intellectual Property Rights and Innovation: Evidence from the Human Genome," *Journal of Political Economy*, 121, 1–27.

——— (2015): "Intellectual Property Rights and Innovation: Evidence from Health Care Markets," in *Innovation Policy and the Economy*, ed. by J. Lerner and S. Stern, University of Chicago Press, vol. 16, 53–87.

——— (2017): "How Do Patents Affect Research Investments?" *Annual Reviews of Economics*, 9, 441–469.

BIBLIOGRAPHY

WISSENSCHAFTSKONFERENZ, G. (2014): *Chancengleichheit in Wissenschaft und Forschung: Vierzehnte Fortschreibung des Datenmaterials (2012/2013) zu Frauen in Hochschulen und außerhochschulischen Forschungseinrichtungen*, Gemeinsame Wissenschaftskonferenz GWK, http://www.gwk-bonn.de/fileadmin/Papers/GWK-Heft-40-Chancengleichheit.pdf.

WISWALL, M. (2013): "The Dynamics of Teacher Quality," *Journal of Public Economics*, 100, 61–78.

WOOLDRIDGE, J. M. (1999): "Distribution-free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, 90, 77–97.

WU, T. (2012): "Taking Innovation Seriously: Antitrust Enforcement if Innovation Mattered Most," *Antitrust Law Journal*, 78, 313–328.

WUCHTY, S., B. F. JONES, AND B. UZZI (2007): "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 316, 1036–1039.

ZABALZA, A. (1979): "The Determinants of Teacher Supply," *Review of Economic Studies*, 46, 131–147.