

Furman, Jeffrey L.; Nagler, Markus; Watzinger, Martin

**Working Paper**

## Disclosure and subsequent innovation: Evidence from the patent depository library program

Discussion Paper, No. 136

**Provided in Cooperation with:**

University of Munich (LMU) and Humboldt University Berlin, Collaborative Research Center Transregio 190: Rationality and Competition

*Suggested Citation:* Furman, Jeffrey L.; Nagler, Markus; Watzinger, Martin (2019) : Disclosure and subsequent innovation: Evidence from the patent depository library program, Discussion Paper, No. 136, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin

This Version is available at:

<https://hdl.handle.net/10419/194032>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

---

# Disclosure and Subsequent Innovation: Evidence From the Patent Depository Library Program

---

**Jeffrey L. Furman** (Boston University Questrom School of Business)  
**Markus Nagler** (LMU Munich)  
**Martin Watzinger** (LMU Munich)

Discussion Paper No. 136

January 8, 2019

# Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program

Jeffrey L. Furman, Markus Nagler and Martin Watzinger\*

October 30, 2018

How important is information disclosure through patents for subsequent innovation? To answer this question, we examine the expansion of the USPTO Patent Library system after 1975. Before the Internet, patent libraries gave inventors access to patent documents. We find that after patent library opening, local patenting increases by 17% relative to control regions. Additional analyses suggest that the disclosure of technical information is the mechanism underlying this effect: inventors start to cite more distant prior art and the effect ceases after the introduction of the Internet. Our analyses thus provide evidence that disclosure plays an important role in cumulative innovation.

---

\*Jeff Furman: Boston University Questrom School of Business and NBER, 595 Commonwealth Ave. Boston, MA 02215, USA; Markus Nagler and Martin Watzinger: LMU Munich, Department of Economics, Akademiestrasse 1, 80799 Munich, Germany; Furman (corresponding author): furman@bu.edu. We thank Sharon Belenzon, Ben Jones, Abhishek Nagaraj, Ariel Dora Stern, Scott Stern, Toby Stuart, Rosemarie Ziedonis, Bo Zhao, and participants at the NBER-AIEA Hong Kong, NBER-PIE program meeting, NBER Productivity Seminar, REER Atlanta, TILEC Tilburg, and UT-Austin Empirical Patent Law conferences, and numerous seminars for helpful comments and suggestions. We thank Regina Seibel for excellent research assistance. We also thank Christopher Dillon, Eileen Fischlschweiger, and Larayne Dallas for their helpful and patient replies to all our questions. Markus Nagler and Martin Watzinger gratefully acknowledge financial support from the DFG through CRC TR 190. Nagler gratefully acknowledges financial support by the Elite Network of Bavaria through Evidence-Based Economics and by the DAAD through a scholarship for doctoral students. Jeff Furman gratefully acknowledges financial support from NSF SciSIP grant, SES-1564368. Parts of this paper were written while Markus Nagler was visiting MIT Economics and while Martin Watzinger was visiting the Questrom School of Business at Boston University. We thank both departments for their hospitality.

*Patent law requires disclosure for the same reason that innovators dislike it: it is the vehicle by which technical knowledge is passed from the patenting firm to its competitors.*

Scotchmer (1991)

## 1 Introduction

The disclosure of technical information is one of the patent system’s central economic functions (e.g., Machlup and Penrose 1950; Scotchmer and Green 1990; Scotchmer 1991; Romer 1990). In legal debate, the U.S. Supreme Court has labeled disclosure the, “quid pro quo of the right to exclude.”<sup>1</sup> In economics, it has been described as a potential microfoundation for the knowledge spillovers that drive economic growth (Romer, 1990). Ideally, patent disclosure will facilitate follow-on innovation by transmitting useful knowledge and by avoiding unnecessary duplication of investment in innovation. In practice, however, intellectual property lawyers have expressed doubts about whether patent disclosure is, indeed, effective in fostering cumulative innovation (e.g. Roin, 2005; Lemley, 2012). Skeptical scholars argue that strategic behavior in crafting patent documents can make it difficult for follow-on inventors to extract key information from prior art searches and that many inventors do not even read patents because this increases the legal risk of “willful infringement.” As a result, such researchers are doubtful that, in practice, patent disclosures transmit truly valuable information to potential future inventors.

Answering the question of how important information disclosure through patents is for subsequent innovation is of first order importance for the design of the patent system and for our understanding of how to increase cumulative innovation. Empirical evidence on this question is scarce, however, because of a fundamental challenge for causal analysis: The patent system makes the right to exclude competitors dependent on disclosing technical information. This leaves little variation to measure the “enablement effect” of disclosure, i.e., the value of information provision on subsequent innovation separately from the effects of exclusion (Graham and Hegde, 2015; Hegde and Luo, 2017; Williams, 2017; Gross, 2018; Baruffaldi and Simeth, 2018).

In this paper, we analyze the large scale expansion of the USPTO Patent and Trademark Depository Library (PDL or patent libraries) system from 1975 to 1997 to investigate the effect of disclosure of patent information on regional innovation. Before the internet, such patent libraries were the only places outside of USPTO headquarters in Washington DC that provided public access to the full range of technical information available on patent

---

<sup>1</sup>Kewanee Oil Co. v. Bicon Corp., 416 U.S. 470 (1974).

documents and that provided tools to search for prior art. Opening a patent depository library in a particular region may, therefore, have reduced by a substantial degree the costs local inventors had to bear to access prior art. While exclusion rights remain national (and constant across region), the opening of patent libraries yields effective variation across regions in the extent of ‘patent disclosure’ during the pre-Internet era.

With the aim of information diffusion in mind, the patent library system was founded in the 1870s to provide patents and innovation-related resources for inventors, entrepreneurs, and incumbent firms. By 1975, 20 libraries had been established, primarily in New England and East of the Mississippi. Beginning in 1975, the USPTO embarked on an effort to open at least one patent library in each of the U.S. states to increase the percentage of U.S. citizens with a patent collection in their commuting zone. This goal was achieved in 1997. We focus our analysis on this period of library system expansion. Although we refer to patent library “opening,” establishing Patent Depository Libraries did not require the construction of new facilities. Instead, opening required that existing libraries dedicated sufficient space, staff, and resources for patent library materials and received official designation as a USPTO Patent Depository Library.

To estimate the impact of opening up a patent library on regional innovation, we compare the change in the number of ultimately-granted patents (by year of application) filed in the region proximate to the newly-opened library after opening with the change in the number of patents in the regions around a matched control sample of Federal Depository Libraries (FDLs). The 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, i.e., providing the public with official documents, nearly all patent libraries are also Federal Depository Libraries. 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.” Patent libraries typically served initially as FDLs and only later became patent libraries, making FDLs in the same state a natural control group.

FDLs are a control group that will provide a valid counterfactual if, in the absence of the opening of the patent library, the number of patents in the region would have followed the same trend as those of the regions around the control libraries. One potential concern about this identification approach is that in the period from 1975 to 1997, libraries could apply to become Patent Depository Libraries. As a consequence, our results could be spurious if the local librarian or the USPTO based their decisions regarding library opening on expected future patenting. While we cannot exclude this threat completely, we document that this is only a concern if the librarian submitting the application (or the USPTO) is able to correctly predict the exact year and the exact place of an increase in future patenting of start-ups

that are active in technologies where patent disclosure is important.

In our main specification, we find that the number of patents within 15 miles of newly-opened patent libraries increased by 17%, an average of around 2.5 patents per 100,000 persons per year. We do not find a negative effect on patent quality, which suggests that the additional patents induced by PDL opening are not of lesser economic value than those produced prior to library opening. This effect of library opening is, however, highly localized and becomes insignificant outside of the typical library commuting distance of 15 miles. Consistent with the prospect that improved access to patent technical information is driving the main effect, we find that patenting increases to a greater degree among young companies, which plausibly face larger barriers to access patents than did larger enterprises during this period. The increase in patenting is most pronounced among patent libraries that are also university libraries, implying a complementarity between access to patent knowledge and technical education for the production of innovation.

We demonstrate that it is unlikely that concurrent shocks drive these effects. In the years before library opening, the number of patents per capita are similar in the regions around the control and to-be-treated libraries. This is consistent with the parallel trends assumption of differences-in-differences analyses. There is also no differential trend between control libraries, suggesting that the libraries do not simply relocate innovative activities from nearby regions. We also show that our results are robust to alternative specifications, for example to including patent library-specific trends, to only using solo or co-located inventors, to using alternative distance circles around libraries, and to using patents counts as the dependent variable while controlling for (time-varying) population.

Our results are robust when we use regions of future, not-yet-opened, patent libraries as an alternative control group. In this specification, young and small companies benefit substantially more from patent library openings than old and large firms, in line with our main results and in line with the notion that libraries changed access to prior patents for resource-constrained inventors the most. We can also construct a “synthetic” doppelgaenger for each patent library by computing the number of patents around patent libraries had these kept their pre-opening shares among all U.S. patents constant. We again find a strong increase in patenting after the opening of the library relative to this synthetic control group. All control groups show very similar trends and our results are similar irrespective of our choice of control group.<sup>2</sup>

In additional analyses, we find three pieces of evidence that the mechanism driving the effect is most likely improved access to patented technical information. First, the effect is most pronounced in chemical technologies, for which patent disclosure matters most for follow-on

---

<sup>2</sup>In Appendix B.2 we show the average number of patents for the three different control groups over time.

innovation. This is consistent with prior survey research that documents the importance of patenting for these technologies, in which patents report valuable and specific knowledge that is, indeed, read by follow-on innovators (Mansfield, 1986; Levin *et al.*, 1987; Cohen *et al.*, 2000; Gambardella *et al.*, 2011). Again, the effect for chemical patents arises only *after* patent library opening. Second, inventors start to cite more geographically and more technologically distant prior art after a patent library is opened. After a library opening, inventors thus start to use knowledge that is less local and the geography of innovation becomes more dispersed. This suggests that patent libraries facilitate the recombination of ideas across fields and contribute to like-minded inventors building on each others' ideas. Third, we find that the effect is strongest after patent libraries introduced computer databases to search for prior art and that the effect vanishes after the internet made patent literature universally available and readily-searchable. This result suggests that simple trends are not driving the results and is consistent with the prospect that access to prior art is a key factor in explaining patenting patterns proximate to patent libraries.

To ascertain whether library opening affects economic outcomes other than purely patenting, we examine data from the U.S. Census on regional business dynamics. We find that Metropolitan Statistical Areas (MSAs) in which patent libraries were established experienced significantly greater rates of small firm entry and exit than did MSAs with comparable FDLs that did not receive such libraries. In addition, local job creation accelerated in regions with patent library openings and this effect was driven by new entrants. These results suggest that the impact of PDLs is not simply an artifact of increased patenting, but is consistent with the prospect that patent libraries affected local entrepreneurial environments.

Our study demonstrates that patent disclosure contributes to subsequent innovation and that this effect appears only if patents are informative. We therefore contribute to the discussion on the benefits of patent disclosure and the merits of the patent system as a whole (Williams, 2017). Our evidence is consistent with the argument that a patent, “serves to disseminate technological information, and that this accelerates the growth of productivity in the economy” (Machlup, 1958, p.76).<sup>3</sup> Not only can disclosure facilitate the market for ideas (Hegde and Luo, 2017), but we find that it can actually help inventors in producing new ideas. By finding particularly strong effects in chemical innovations but weaker in other areas, a technological area in which disclosure is thought to be particularly effective, our study also offers qualified support for critics that argue that the usefulness of disclosure through patents for subsequent innovation is limited (e.g., Roin, 2005; Lemley, 2012; Chien,

---

<sup>3</sup>In a similar vein, Romer (1990, p.84) writes that patent disclosure increases economic growth because, “other inventors are free to spend time studying the patent application for the widget and learn knowledge that helps in the design of a widget.”

2016) and that the benefits of reading patents are mixed (Arora *et al.*, 2008; Gambardella *et al.*, 2011; Cohen *et al.*, 2002; Hall and Harhoff, 2012; Ouellette, 2012, 2017).<sup>4</sup> Normatively speaking, our empirical analysis shows the potential upside of more stringent disclosure for innovation (Bessen and Meurer, 2008; Fromer, 2008).

More generally, our study contributes to the literature on research-enhancing institutions by showing that investments in patent libraries helped to fuel regional innovation. Research enhancing institutions lower the cost of access to useful knowledge and thus help to foster geographical and intertemporal spillovers on which economic growth is based (Mokyr, 2002). For example, Furman and Stern (2011) demonstrate that biological resource centers, libraries of living organisms, can foster follow-on innovation by providing open and low cost access to life sciences research materials. In recent work, Biasi and Moser (2016) show that reducing the access costs to science books during World War I increased scientific output particularly in those regions in which libraries bought these books. Our research contributes to this literature by showing that patent libraries increased innovation across U.S. states by improving access to patent documents.

Historical analyses of the U.S. patent system have noted its role in democratizing innovation, i.e., in enabling innovation to take root in various geographic regions, across socioeconomic groups, and among different types of enterprises (Machlup, 1958; Scotchmer and Green, 1990; Lamoreaux and Sokoloff, 1999; Khan, 2005). Our results provide evidence that, though its 1975-1997 Patent Depository Library program, the USPTO continued to play a role in spreading innovation across regions and organizational types. In an age before the internet, searching patent documents at a close-by patent library provided an accessible way to study such prior art, especially for resource constrained inventors. For example, avid patent library user and hearing aid innovator, Geoffrey Ball, stresses in his autobiography the importance of technical information in prior patents to his own work, in which he lauded the Sunnyvale CA patent library as the “only place to research patents” available to him when he was initially researching his inventions (Ball, 2012).

The remainder of this paper is organized as follows: Section 2 describes the U.S. Patent Depository Library Program and the Federal Depository Library Program. In Section 3 we describe the data and the empirical strategy. In Section 4, we show that opening a patent library increased innovation in its close vicinity and present robustness checks. In Section 5, we present evidence on the underlying mechanism. Section 6 presents evidence on the impacts of patent libraries on small businesses. Section 7 concludes.

---

<sup>4</sup>Newer studies on the American Inventor Protection Act show that many inventors voluntarily disclose their inventions, leading to earlier licensing deals (Graham and Hegde, 2015). Current work by Gross (2018) documents several effects suggesting that technologies whose disclosure was inhibited by the U.S. World War II patent secrecy program experienced a decrement to follow-on innovation.



## 2 The U.S. Patent Depository Library Program<sup>5</sup>

### The Establishment of the Patent Deposit Library Program

In the years following the Civil War, the U.S. Congress acknowledged that the increasingly industrial and innovation-focused country could benefit from expanding access to technical information contained in patent documents. In the early 1870s, Congress enabled the creation of a nationwide network of Patent and Trademark Depository Libraries. Prior to 1871, official patent documents were housed and available for widespread perusal only at the Patent Office in Washington DC. In that year, federal statute 35 USC 12 officially entitled The Patent Office to distribute copies of patents to designated libraries outside the capital. In addition, the Patent Office began in 1872 to publish and disseminate weekly the *Official Gazette*, which reported a brief abstract and a representative drawing of each patented invention.<sup>6</sup> These efforts were particularly important in stimulating the spread of technical information at a time when the cost of traveling to Washington DC effectively limited access to patent prior art. According to Lamoreaux and Sokoloff (1999), patent agents and solicitors emerged in the 1800s as an important institution that inventors and firms outside of Washington could use in order to obtain information about and build upon new inventions. The patent depository libraries were another key institution aiming to support innovation via information diffusion.

The first set of patent depository libraries were established in the 1870s at 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. New libraries were slowly added over the next few decades and, by 1975, the number of patent libraries had grown to 20, most of which were located in the industrial Midwest and eastern seaboard. By the 1970s, each library received weekly shipments of unbound paper patents, the *Official Gazette* of the U.S. Patent and Trademark Office, and two search indices.

---

<sup>5</sup>The history section follows the descriptions in Sneed (1998) and Jenda (2005).

<sup>6</sup>The *Gazette* was one of a number of publications that provided limited information about patented inventions. For example, the journal *Chemical Abstracts* began publishing abstracts of chemical patents in 1907 and the periodical *Scientific American*, which began publication in 1845, featured patent summaries throughout its history. Like the *Gazette*, however, each of these sources published only patent abstracts and up to one drawing and, did not, therefore, provide the rich source of technical information available in original patent documents or in patent depository libraries.

## The Expansion of USPTO Patent Library System 1975-1997

Because access to patent documents remained limited to paper-based methods, individual inventors and small and medium-sized enterprises engaged with patent prior art in the early 1970s in much the same way as they had in the 1870s, i.e., through intermediaries, including patent agents, or via travel to locations with complete patent records. Recognizing the need for expanded facilities after one hundred years of relative inactivity, the USPTO began an aggressive expansion of the patent library system beginning in 1975.<sup>7</sup> The revived program established goals of increasing the number of patent libraries by at least three per year and, ultimately, operating at least one patent library in each state. This latter aim was achieved in 1997. The map in Figure 1a identifies the twenty libraries in operation before 1975, while Figure 1b lists all patent libraries opened after 1975. Figure 2 shows the expansion of the patent depository library system over time.<sup>8</sup> Currently, about half of the PDLs are based in academic libraries and nearly as many are affiliated with public libraries.<sup>9</sup> After 1997, the patent library system adopted a new goal of controlled growth in areas with high population combined and high patent and trademark activity (Sneed, 2000).

### How (and why) Libraries Joined the Patent Depository Library Program

Beginning in 1975, existing library facilities became eligible to apply to become PDLs if they fulfilled a set of requirements. First, libraries had to demonstrate that they had the physical capacity (space) to acquire and make available for use a collection of *all* U.S. utility patents issued twenty years prior to the date of library designation (i.e., *all* patents in force). Second, each patent library had to commit to employing and training sufficient staff to assist the public in the search for prior art. To ensure adequate training, each patent library had to send a representative to the annual PDL Training Seminar in Washington DC.<sup>10</sup> Third, they had to provide free public access and a collection of search tools for the public. According to the USPTO, the first library in each state that (a) applied for PDL status and (b) successfully fulfilled these criteria would receive a designation as a patent depository library.

These criteria implied that larger libraries, such as university libraries and city public libraries, were able to fulfill the resource requirements of becoming a patent depository library. Over time, however, the space requirement became less a concern after the intro-

---

<sup>7</sup>This effort was initiated by USPTO Assistant Commissioner William I. Merkin, beginning with an assessment of the patent library system in 1974.

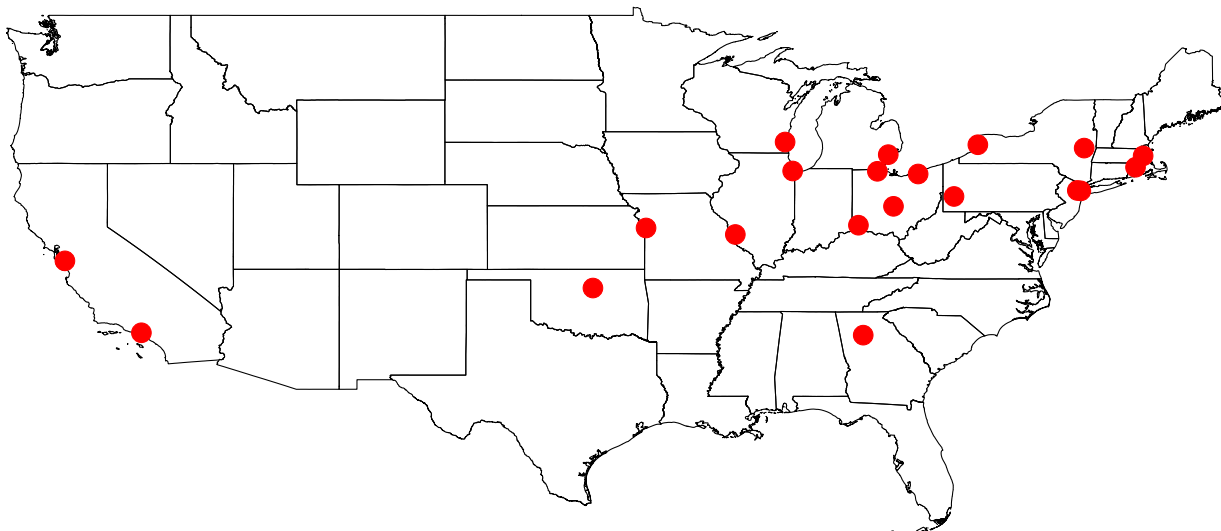
<sup>8</sup>Table A-1 and Table A-2 in the appendix list patent libraries up to 2002.

<sup>9</sup>Since 1871, six PDLs 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 required services.

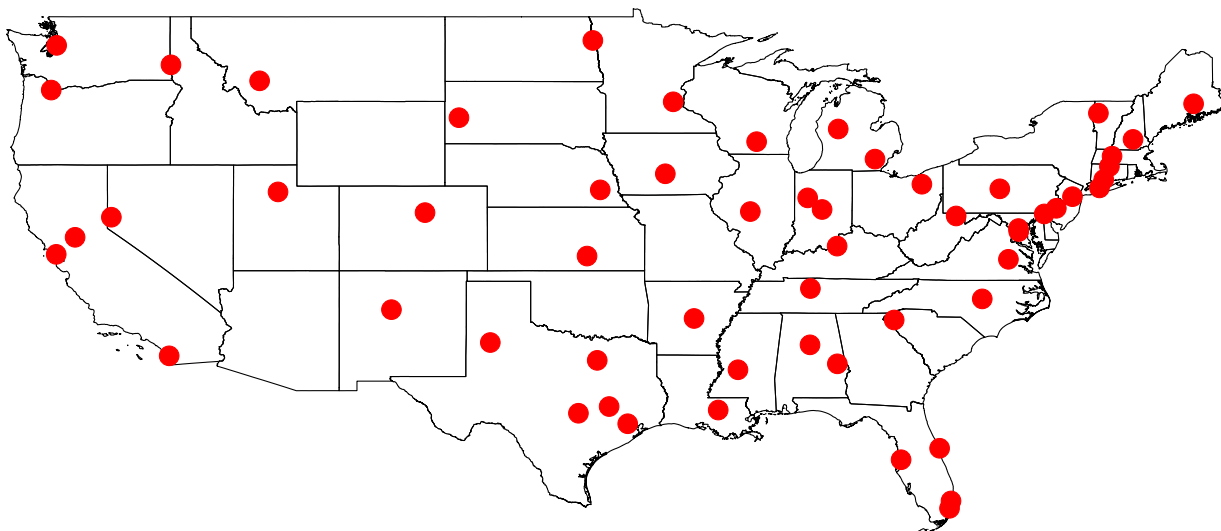
<sup>10</sup>Indeed, several of the librarians we interviewed mentioned that the opportunity to participate in the annual training was a nontrivial reason for their association with the Patent Deposit Library program.

Figure 1: Location of all Patent Libraries in the U.S.

(a) Opened from 1870 to 1974

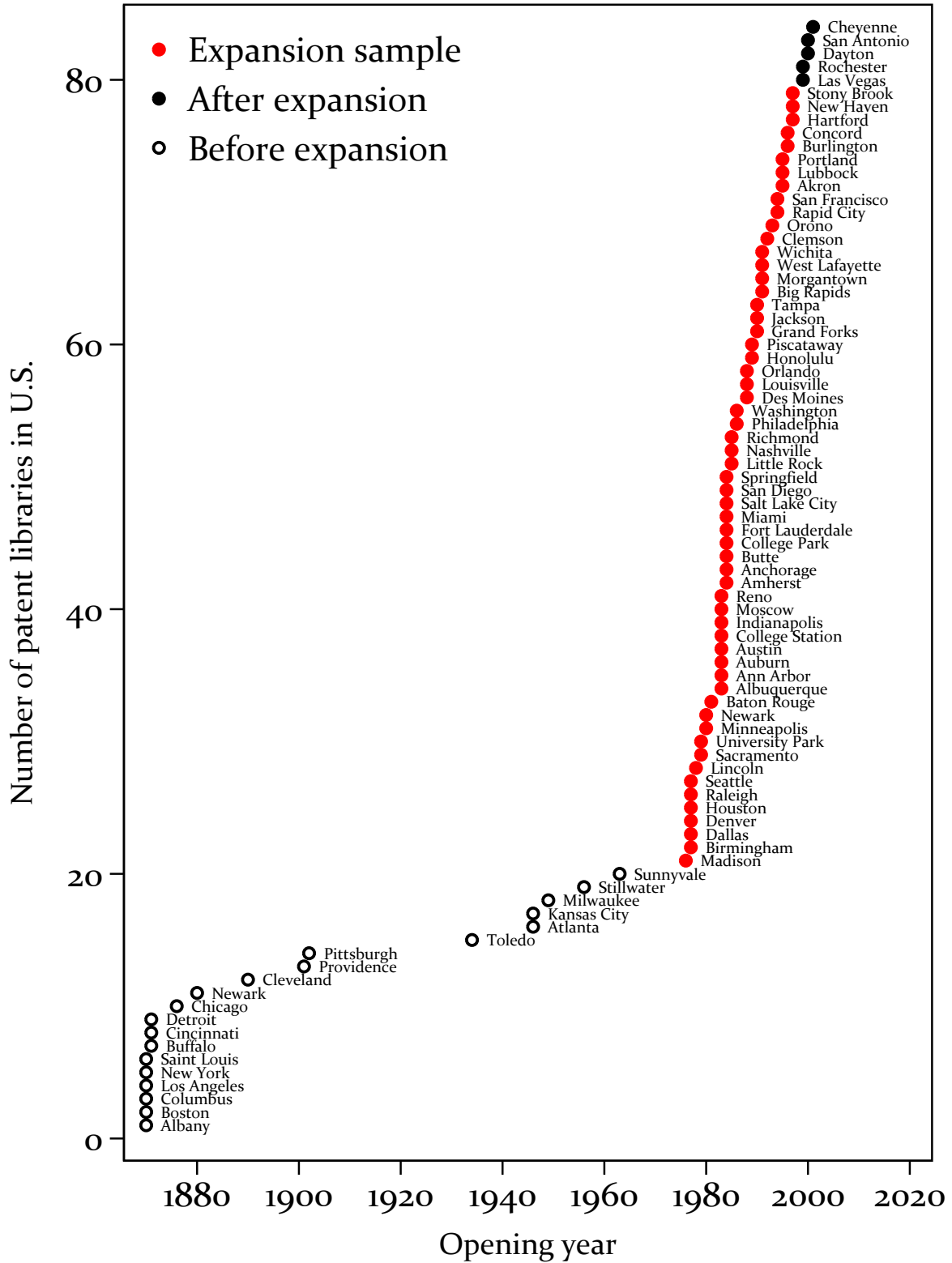


(b) Opened from 1975 to today



**Note:** Figure 1a shows the position of patent libraries in the continental United States opened before the major expansion from 1975 to 1997. Figure 1b indicates the location of patent libraries opened after 1975.

Figure 2: The Expansion of the Patent Depository Library Program



duction of microfilm. Indeed, the conversion from paper to microfilm distribution has been cited as a reason why many new libraries joined the program after 1982. The fact that the process of becoming a patent library during the 1975-1997 period was initiated by the library itself rather than solicited by the USPTO, may explain, in part, why patent libraries were not opened in the sequence one might expect *a priori*. For example, Honolulu HI and Big Rapids MI each received patent libraries before either New Haven CT or San Francisco CA (Figure 2). The librarians that we interviewed reported a number of reasons that their institutions applied to join the PDL community during this period. Some librarians we interviewed suggested that their libraries applied to join as a result of their institutional missions or out of a sense of duty to their patrons. These factors may reflect local demand for patent information. However, interviewees also mentioned factors more idiosyncratic and less predictable in driving their library's participation, including the perceived attractiveness of annual PDL trainings in Washington DC and the professional benefits of participating in the PDL librarian community.<sup>11</sup>

Most patent libraries had prior experience handling government documents as Federal Depository Libraries before applying to become patent depository libraries. 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 U.S. Congressional Districts.<sup>12</sup> Because of this structure and the requirements associated with serving as in either library program, we believe that Federal Depository Libraries constitute a natural control group for patent depository libraries.<sup>13</sup>

### **What Services did Patent Depository Libraries Provide (and to whom)?**

The main aim of the patent deposit libraries, both in the modern era and in the 1800s, was to provide access to technical information to potential users and to help them with prior art searches. Throughout the program's history, however, patent librarians have been embarrassed from providing legal advice or other legal services. Thus, their services have focused

---

<sup>11</sup>The annual PDL trainings in Washington DC appear to have been both highly valued professionally and personally enjoyable to the PDL librarian community. Both the professional lessons and personal reflections are documented in the Patent and Trademark Resource Center Association Newsletters, which are available for review at <http://ptrca.org/newsletters>.

<sup>12</sup>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, all libraries at land-grant colleges and universities, libraries of federal agencies, the highest appellate court of a state, and accredited law schools automatically qualify for the status of Federal Depository Library.

<sup>13</sup>The USPTO continues to operate the patent library program even after the advent of freely available patent document search engines, like Google Patents. The librarians we interviewed suggested that the current libraries, now called Patent and Trademark Resource Centers, aim to create value for the communities they serve by assisting with the search for prior art and by helping users negotiate databases that offer more sophisticated prior art search capabilities than publicly-available resources.

on information provision. The records of the annual conference document the exceptional dedication of the library professionals to these tasks (see., e.g., Sneed, 1998, and Oliver, 2002). Surveys of patent depository library users in the 1990s suggest that the libraries served mainly local users and inventors in particular. Specifically, the 1991/1992, 1997, and 2002 surveys suggests that the majority of users were inventors or entrepreneurs, while fewer than 16% were attorneys, patent agents, or other legal staffers. The 1997 and 2002 surveys report that the median user estimated traveling between 11-20 miles to use the library, with 49% of 2002 users and 38% of 1997 users traveling fewer than ten miles (Brown and Arshem, 1993; Patent and Office, 1999).

### 3 Empirical Setup

#### Identification: Federal Depository Libraries as Control Group

To measure the impact of opening of a patent library on regional innovation, we need a counterfactual estimate of what would have happened to patenting in region if the library had not opened. To do this, we develop a control group that includes regions that are geographically proximate to newly-opened patent depository libraries, that have medium or large Federal Depository Libraries (FDLs), and that are within the same state as the treated patent library.<sup>14</sup>

Regions around Federal Depository Libraries are particularly attractive as control group for three reasons: First, FDLs already handle government documents, which is one of the key criteria for becoming a patent library. Second, there are at least two FDLs in each congressional district, which ensures that FDLs would be able to serve as control group in each state. Third, medium and large FDLs satisfy the formal requirement for becoming patent libraries. They are the likely to possess the space, human capital, and library infrastructure required to become patent libraries. Indeed, 82% (53 of 64) of new patent libraries after 1975 are FDLs.

In our main specification, we focus on patent libraries that are also Federal Depository Libraries, as this enables us to work with a well-matched control group. In our estimation sample, we drop (a) all patent libraries that were opened before 1975 or after 1997, as the motivation for opening libraries in those periods was different than it was during the period of dedicated program expansion of the , (b) all patent libraries that are not Federal Depository Libraries, and (c) all PDL libraries for which we cannot identify an FDL control library

---

<sup>14</sup>The Federal Depository Library program classifies libraries as small if they contain fewer than 250,000 volumes in the library, medium if they contain 250,000 to one million volumes, and large if they possess more than one million volumes of public materials.

Table 1: Sample Selection

Sample	Patent Libs	FDLs	Libraries dropped
All patent libraries	84	661	
Only Patent Libraries after 1975 and before 1997	59	493	
Only Patent Libraries that are also FDLs	49	408	Anchorage (AK), Concord (NH), Des Moines (IA), Hartford (CT), Jackson (MS), Minneapolis (MN), New Haven (CT), Piscataway (NJ), Springfield (IL), Washington (DC)
Estimation sample	48	406	Burlington (VT)

between 15 miles and 250 miles. We also drop the library in Burlington VT, because we cannot find a suitable control region within its state. Burlington hosted a primary research facility of IBM during the sample period and its patents per capita ratio vastly exceeded that of other regions. Over the sample period, the Burlington library region averaged more than 295 patents per 100,000 persons, while sample average was just less than fifteen and the region with the second highest patents per capita was less than one hundred (Newark DE, home of a Dupont primary research facility). Our primary estimation sample, thus, includes 48 patent libraries that opened after 1975, along with 406 control libraries. In Table 1 we show how each of these selection steps influences the sample composition. Figure 3a shows the position of all patent libraries and all Federal Depository Libraries in our sample.

FDLs will be a valid control group if the number of patents in patent library regions would have followed the same trend as the number of patents around the control libraries were patent libraries not opened. Our identification assumption would be threatened if librarians or administrators had applied to become Patent Depository Libraries (or were selected by the USPTO) in the expectation that innovative activities were about to burgeon in their regions. Although we cannot rule out this possibility, three factors suggest that this may not be a substantial threat to causal interpretation. First, to condition their application for becoming a patent library on future innovation, local librarians must be able to predict accurately near-term changes in local private sector innovation. While librarians likely have local insights, they would need to anticipate coming boosts in local patenting that do not involve pre-trends, compile their applications, and have them approved at precisely the time that local patenting is about to increase. Considering the mix of idiosyncratic reasons that

librarians report as having played important roles in libraries' application decisions, we consider this unlikely. Second, the program expansion from 1975 to 1997 aimed at opening one library in each state, with the goal of supporting equal access to patent materials across the country. Thus, it seems less likely that the USPTO accepted library applications based on changes in expected future patenting. Indeed, according to the USPTO, patent library status was supposed to be conferred upon the first qualified library in each state that applied for program participation. After 1997, however, the USPTO switched to favoring regions with high patenting per capita. Third, while most centers of innovation ultimately receive patent libraries, the key to our identification strategy is that the timing with which the libraries are opened must be random with respect to innovation trends. The particular dates on which regions receive libraries does not follow a pattern of increasing or decreasing innovation importance, either in levels or in changes.

To explore whether the assumptions underlying our identification strategy are reasonable, we conduct several robustness checks in Subsection 4.2. First, we show that before the patent library was opened, the trend in the number of filed patents per capita was the same around the soon to be designated patent library and the control libraries. This speaks in favor of (counterfactual) parallel trends. Second, we find little effect if we assign pseudo treatments to the closest control library. This speaks in favor of the stable unit treatment value (SUTVA) assumption. Third, we use a host of different specifications for the control group and find that our results are robust. Most importantly, we report the results of a model that identifies the impact of patent library opening using regions around not yet opened patent libraries as controls for patent libraries that were opened earlier.

## Data Sources

For our empirical analysis, we combine data on libraries with geolocated patent data and population data from the U.S. Census. The data on the opening dates of each patent library is from Jenda (2005) and the complete list of Federal Depository Libraries is from the online Federal Depository Library Directory.<sup>15</sup> We obtain Patent data from PATSTAT. To identify the geographic location of the inventors and inventor disambiguation we rely on the data of Balsmeier *et al.* (2017) and of Morrison *et al.* (2017). If there are several inventors on a patent, we allocate each location a share of the patent.<sup>16</sup> Appendix A gives a complete list

---

<sup>15</sup>The Federal Depository Library Directory is available on <https://catalog.gpo.gov/fdlpdir/FDLPdir.jsp> (last accessed 2017-07-30).

<sup>16</sup>Using patents as an indicator for innovative output is standard but not uncontroversial. In our particular case, patent libraries also might increase patenting without increasing innovation because they might make it easier to file a patent or because the librarians might 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



of patent libraries with opening dates and provides a step-by-step description of the data processing.

Figure 3b plots the patent libraries together with patent data across space. To aid visualization we also plot the centroid of each county in light gray. Places with more patents have larger dots. There is an apparent correlation between the location of patent libraries and the number of patents. Yet, places with many patents are also places with a larger population and thus more potential inventors. Thus, to adjust for different city sizes, we normalize the number of patents by the population in the area. For population data, we use the U.S. Census data for incorporated places at the end of the sample period in 2010.<sup>17</sup>

Table 2 shows summary statistics for patent libraries and matched Federal Depository Libraries in the year before the opening of the patent library in levels. Some of the means of the treated observations are close to being statistically significantly different from those of the control group. This is due to some outlier regions that do not have patenting activity in some years. In Section B.1 in the Appendix, we document that balancing improves when we drop these observations. All results hold when using this alternative control group.<sup>18</sup>

## 4 Do Patent Libraries Increase Local Innovation?

It is unclear *a priori* whether the opening of a patent library will have an impact on innovation in the library's geographic region. On the one hand, improved access to patent literature could induce local innovation if inventors were to read the patent literature, draw valuable information from it, and, as a result, innovate at lower cost, greater rapidity, or with greater effect than would have been the case in the absence of the library (Machlup and Penrose, 1950; Scotchmer and Green, 1990; Scotchmer, 1991; Landes and Posner, 2003). On the other hand, a number of scholars doubt that increasing the local availability of patent technical information could have such an impact because they expect that effectively-written patents will be opaque and because inventions whose inventive steps cannot be obscured in patent filings can be kept secret rather than shared with competitors via patent disclosures (Levin *et al.*, 1987; Moser, 2011, 2013). Further, legal scholars point out that inventors may wish to avoid reading patents even if they were potentially helpful, because reviewing patent

---

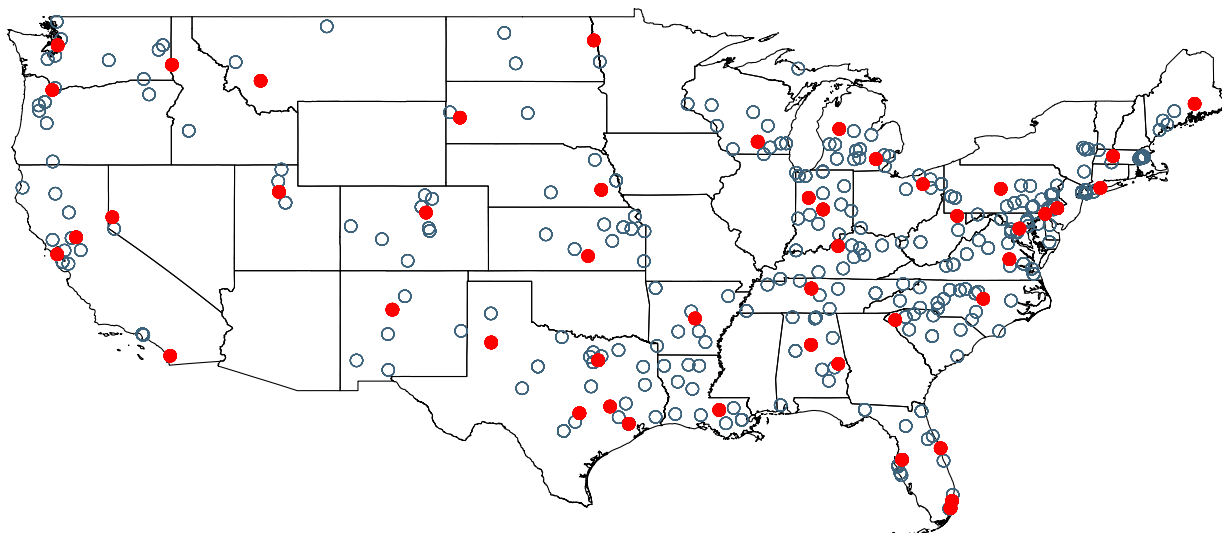
patent libraries are only allowed to help with the search for prior art but not with the preparation of a patent filing.

<sup>17</sup>In our robustness checks, we also use time-varying population. Here, we use the NBER's Census U.S. Intercensal County Population Data, 1970-2014, and compute the population in counties within 15 miles of a library.

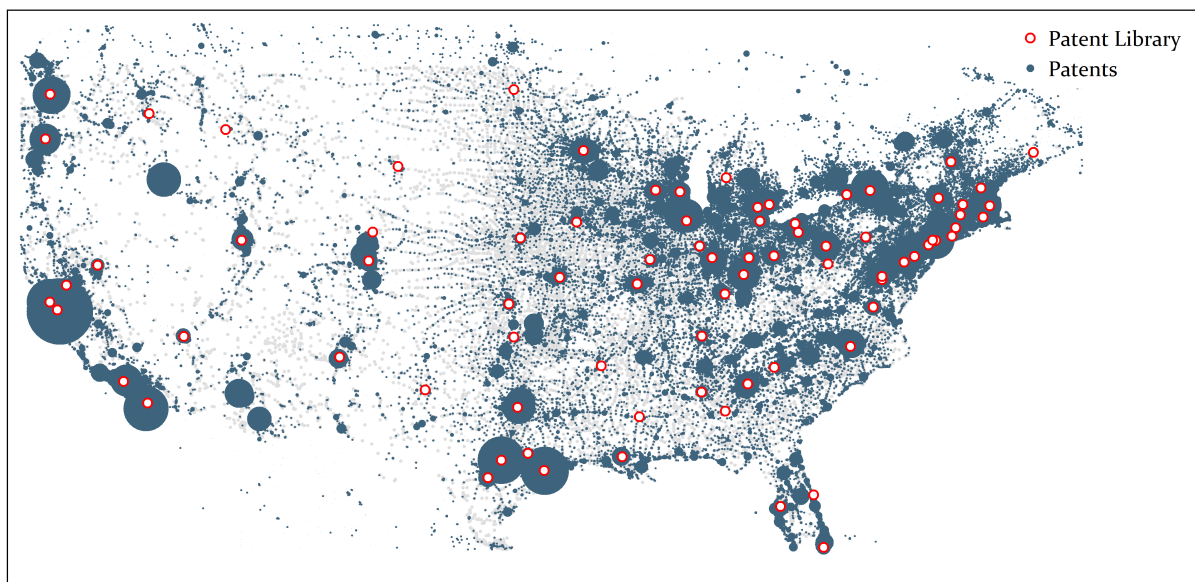
<sup>18</sup>Further, in Table 4 we find that matching libraries by previous patenting or by population does not change our estimates qualitatively and that using different control groups (such as only using the differential timing of PTDL openings within the set of future libraries) yields the same result.

Figure 3: Libraries and Patenting Across Space

(a) Locations of Patent and Control Libraries



(b) Number of Patents and Patent Libraries by Location of Patent



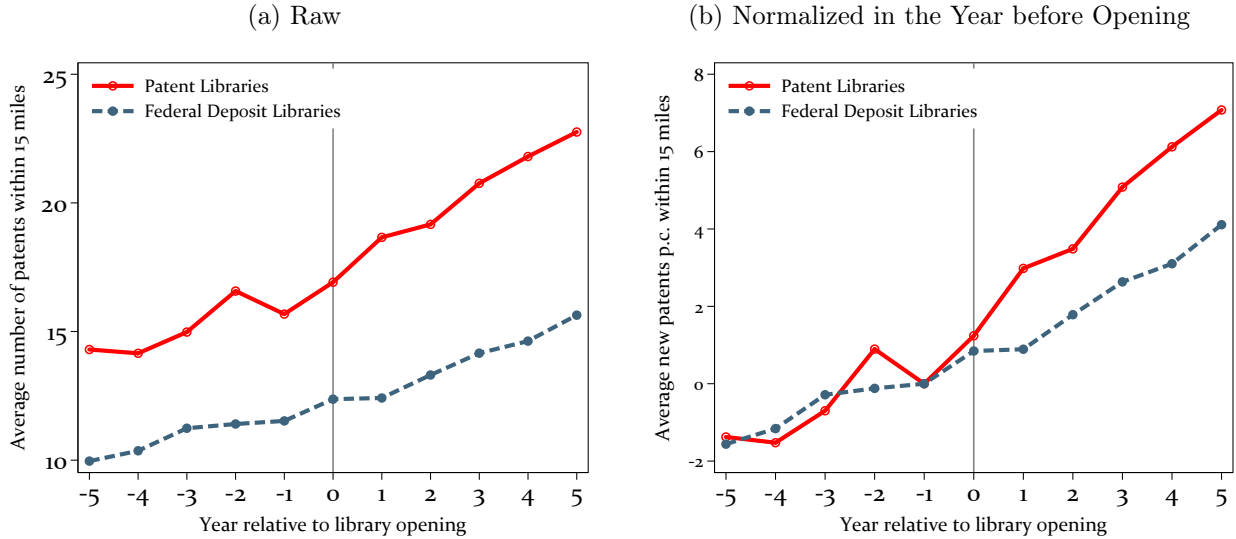
**Note:** In Figure 3a, the red dots show the position of patent libraries. The hollow dots show the positions of control libraries. In Figure 3b, the red dots identify the position of patent libraries. The blue dots show the positions of patents. A larger dot signifies that there are more patents at the same place. To aid visualization we plot the centroids of each U.S. county in light gray.

Table 2: Summary Statistics in the Year before Opening

<i>Main sample</i>				
	Patent Libraries	Control Libraries	Diff	P-Value
Population in 100k	7.60	7.41	-0.19	0.93
Uni Library	0.67	0.69	0.03	0.72
# Patents	128.29	81.82	-46.46	0.09
# Patents/100k	15.68	11.53	-4.15	0.11
Citation-weighted patents	226.76	173.58	-53.18	0.25
Dollar-weighted patents	83.80	115.90	32.10	0.46
# Pat. small firms/100k	7.25	6.03	-1.22	0.26
# Pat. big firms/100k	8.43	5.49	-2.94	0.14
# Pat. young firms/100k	5.45	4.35	-1.09	0.23
# Patents old firms/100k	10.23	7.17	-3.06	0.15
Number of libraries	48	406		
<i>Patents by field</i>				
	Patent Libraries	Control Libraries	Diff	P-Value
Electrical Engineering	2.23	1.94	-0.28	0.60
Instruments	2.27	1.69	-0.58	0.14
Chemistry	4.02	2.01	-2.01	0.11
Process Engineering	2.02	2.14	0.12	0.78
Mechanical Engineering	2.98	2.07	-0.92	0.26
Other Fields	2.14	1.67	-0.47	0.27

**Note:** This table shows the averages of the data for patent libraries and associated control libraries in the year prior to patent library opening. The last two columns show differences with the associated significance levels. Dollar-weighted patents use patent values from Kogan *et al.* (2017) to weight each patent with its dollar value. We windsorize the Dollar values at the 90th percentile to adjust for outliers. A firm is defined as young if its first patent was filed less than three years before the opening of the patent library. Otherwise it is old. A firm is defined as small if it has no more than 20 patents before the opening of the patent library. Otherwise it is large. The p-values result from a t-test with unequal variances.

Figure 4: Compare Averages



**Note:** This figure plots the average number of patents within 15 miles of the patent library (red solid line) and around Federal Depository Libraries (blue dashed line) in the five years before and after the opening of the library. Figure 4a shows the raw average and in Figure 4b we normalize the average relative to its value in the year before the opening.

prior art may expose innovators to willful infringement, which increases their legal exposure to enhanced financial penalties (Roin, 2005; Lee and Cogswell III, 2004). Finally, patent disclosure could also have no or even negative impacts on local innovation if disclosure decreases duplicate inventive efforts. We evaluate these opposing viewpoints empirically in the section that follows, by investigating whether data on regional patenting suggest that patent libraries contributed to local innovation during the 1975-1997 period.

#### 4.1 Primary Analysis: Patenting Increases After Library Opening

We begin by asking whether patent library opening impacts patenting within 15 miles around the new library. In Figure 4a we compare the raw difference in the average number of patents per 100,000 persons around treatment and control libraries. In Figure 4b we subtract from each series its value in the year before the opening of the library to account for different levels. In both cases, the two series begin to diverge in the period after the patent library opened. In Appendix B.2 we show these averages for different potential control groups.

In Figure 5, we plot the yearly difference in the number of newly filed patents per 100,000 population in the 15 miles radius around the control and the patent libraries. For each library region, we consider the five years before and the five years after the library opening and we normalize the number of new patents to zero in the year before opening. We use weights to

adjust for the different number control libraries per patent library to arrive at the average treatment effect on the treated (Iacus *et al.*, 2012).<sup>19</sup>

The data suggest that the number of newly filed patents around the patent library increases significantly after opening. This effect is relatively smaller in the first two years and is larger and statistically significant in each of the third through fifth years after library opening. Prior to the opening of the patent library, the number of patents per capita is similar for treatment and control libraries. This is consistent with the parallel trends assumption and provides some confidence that the estimated coefficient represents a causal effect.

To quantify the size of the effect of opening a patent library, we estimate the following difference-in-differences specification:

$$\frac{\#Patents_{it}}{Population_i} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i \quad (1)$$

where  $i$  indexes each library (i.e., the 15 miles region around each library),  $PatLib_i$  is an indicator equal to one if the library in that region is a patent library, and  $Post_t$  is an indicator equal to one in the years following patent library opening. We incorporate both library and year fixed effects as controls.<sup>20</sup> The coefficient of interest,  $\beta_2$ , measures the average yearly increase in the number of patents around a patent library in the five years after it was opened relative to the period before it was opened and relative to the controls in that period.

We report the results for estimating Equation (1) in Table 3. Column (1) shows that the number of patents per capita (patents per 100,000 persons) in the region around of the patent library increased on average by 2.5 relative to the control group. This implies an increase of 17% relative to the average. This estimate is the primary result in the paper. If we can interpret the regression as causal, it implies that patent library opening induces local innovation in the area proximate to the libraries. The estimate may underestimate the impact of patent disclosure on subsequent innovation, as some patent technical information may have already diffused to these areas through means other than the patent libraries, such as the publication of patent abstracts, the operation of internal patent libraries by large firms, and the ability of firms to access patent information by other means.

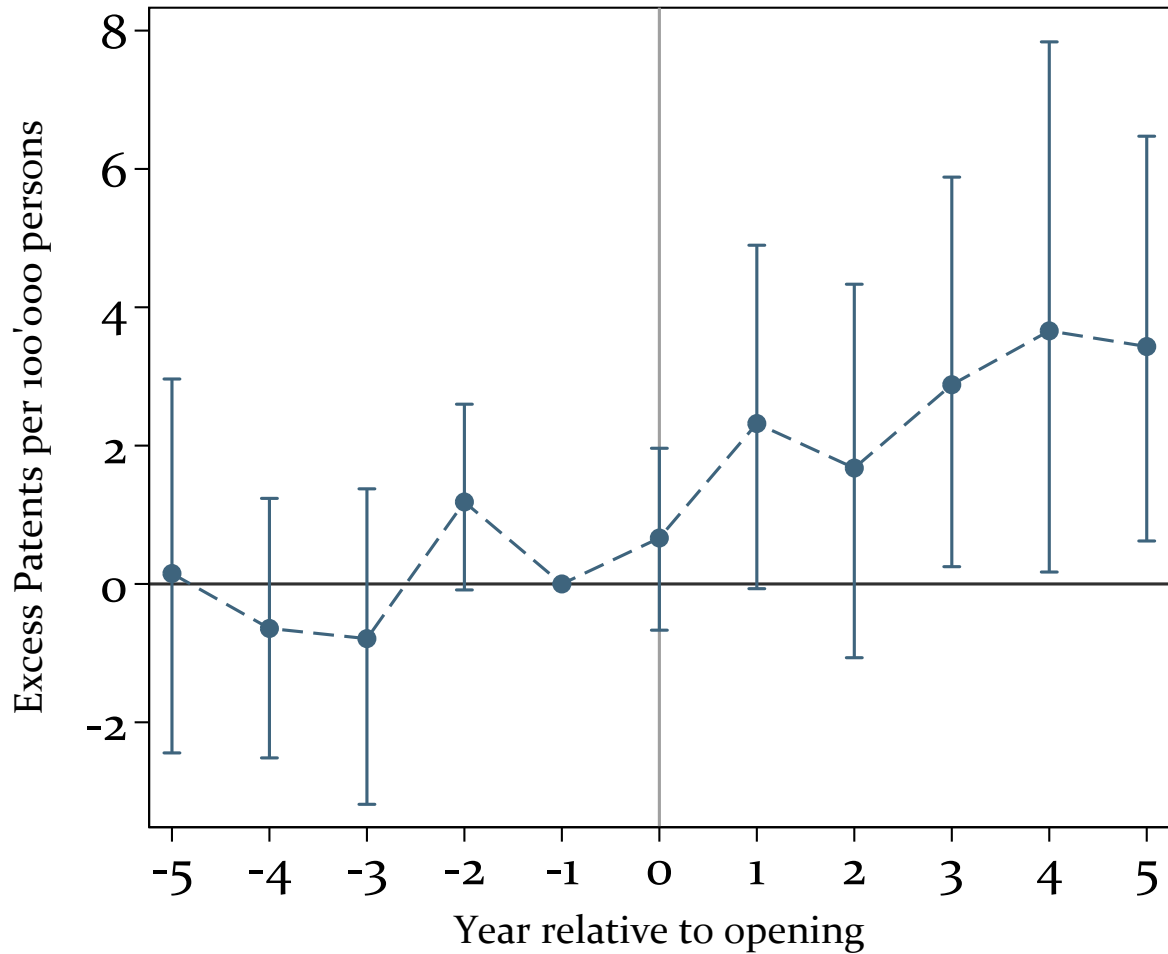
The increase in newly-filed patents we find in (1) does not go hand in hand with a decrease in patent quality. We test this in column (2) by estimating the impact of library opening on citation weighted patents. Using this common approach to account for patent quality, we observe not a decrease in citations, which would be consistent with a decline in quality, but

---

<sup>19</sup>As we show in Table B-2, our results do not depend on using these weights.

<sup>20</sup>The baseline effect of  $PatentLib_i$  is taken up by the library fixed effects.

Figure 5: Non-parametric Evidence



**Note:** This figure shows the yearly average treatment effects on the treated of opening up a patent library on the average number of patents within 15 miles of patent libraries relative to the average number of patents around matched federal depository libraries. The 95% confidence intervals are based on bootstrapped standard errors. We use the weights of Iacus *et al.* (2012) to arrive at the average treatment effect on the treated. We assign each patent library and all Federal Depository Libraries within the same state and within 250 miles as control group. We exclude the patent library of Burlington VT.

Table 3: Patent Libraries and Local Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Patents	Baseline Citations	Dollars	Age		Size		University	
				Young	Old	Small	Large	Yes	No
Post	-0.2 (0.6)	-17.3 (17.2)	-7.8 (8.9)	-0.7** (0.3)	0.4 (0.5)	-0.3 (0.4)	0.1 (0.4)	-0.4 (0.7)	-1.0 (1.1)
Pat Lib x Post	2.5** (1.2)	44.0** (16.7)	42.2*** (15.0)	1.1** (0.5)	1.4 (0.8)	0.9* (0.5)	1.6 (1.0)	3.0* (1.6)	1.1 (1.3)
Mean Dep.	14.5	200.9	93.4	4.9	9.6	7.2	7.3	13.8	15.9
R2 (within)	0.09	0.10	0.14	0.23	0.05	0.12	0.03	0.08	0.17
Obs.	4994	4994	4994	4994	4994	4994	4994	3630	1364

**Note:** 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:

$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within 250 miles as controls. In column (2) we weight each patent with its forward citations. In column (3) we use the patent values from Kogan *et al.* (2017) to weight each patent with its dollar value. We windorize the Dollar values at the 90th percentile to adjust for outliers. In columns (4) and (5) we split the dependent variable by young and old assignees. An assignee is young if it filed its first patent no more than three years before the opening of the library and old otherwise. In the following two columns we split the dependent variable by the size of assignee. An assignee is defined as large if it has more than 20 patents before the opening of the patent library. In column (8) and (9) we consider the subsample where the patent library is also a university library and where it is not. 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 (assigned) patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

rather an increase by 44 citations, which implies an increase of around 20 percent relative to the mean number of citations received. In column (3) we weight each patent by the value of the patent estimated by Kogan *et al.* (2017). The result here implies that the average region experiences a boost in patenting whose value is approximately \$42 million per year accruing to the inventor privately. This compares favorably with the operating costs of a library. For example, the Boston Public Library reported a fiscal year 2015 budget of \$41.6 million, while the San Antonio Public Library and the Free Library of Pittsburgh reported 2015 fiscal year budgets of \$34.9 million and \$48 million, respectively. Considering that patent collections constitute only a small fraction of the total operating expenses of each patent deposit library affords some confidence that the boost in patenting induced by access to patent technical documents is, across the program, justified by the cost. However, note that the costs and benefits accrue to different stakeholders. One key caveat of this analysis is that the Kogan *et al.* (2017) values are based on publicly-held firms, whereas our data include patents issued to all assignee types. In addition, as Kogan *et al.* (2017) note, their measure might be upward biased as a result of their econometric choices.

In columns (4) to (7) we split the dependent variable by the type of assignee. In our analysis, we consider firms to be young if their first patent was filed less than three years prior to the library opening.<sup>21</sup> These young firms may be entrepreneurial ventures, but they may also be existing firms that had not previously applied for patents. The opening of the average patent library increases the number of patents by young companies by 1.1 patents per 100,000 residents, an increase of 22 percent relative to the mean. The effect for old companies, reported in column (5) is larger in magnitude than that of young firms relative to the mean but is noisier and not statistically different from zero. In columns (8) and (9), we split the result by the type of library. We compare the impact of library opening at universities relative to public libraries and find that the impact on innovation is greater among the former. This suggests a potential complementarity between access to prior art and university innovation ecosystems. In Appendix B.3 we demonstrate that the effect is mostly driven by patents assigned to firms and that there is only a small effect associated with patents assigned to universities. There is no clear difference between regions with historically high or low patenting levels, even though the effect is only statistically significant for the latter.

In Table B-2 in the Appendix we also show that our results are robust to not using the CEM weights suggested by (Iacus *et al.*, 2012), to including patent library-specific trends, to only using solo or co-located inventors, to using other circles around the libraries (for example

---

<sup>21</sup>This result is robust to using alternative cutoffs, such as defining young firms as those which file their first patent after patent library opening.



25 or 50 miles), and to using the number and the log number of patents as the dependent variable while controlling for (now time-varying) population. We also show in Appendix B.6 that this effect is robust to controlling for the local number of registered patent attorneys.<sup>22</sup>

## 4.2 Auxiliary Analysis: Increase in Patenting is Likely Causal

The principal concern for our estimation strategy is whether regions that receive a new patent library would have experienced equal boosts in patenting even if they had not received patent depository library facilities. This could occur if the patent librarians were to have accurately anticipated the timing of local innovation bursts and, in this expectation, applied to become a patent library. An analysis of patenting under such circumstances might yield observationally equivalent results, though these additional local patents would not have been induced by the library opening. In this section we report the results from auxiliary analyses that shed light on this possibility. We report the results of these additional analyses in Table 4, comparing new specifications to our preferred specification, which is replicated in column (1). Our conclusion is that such a coincidence is not likely.

### No Effect Before Patent Libraries Opened nor Outside of Local Region

Figure 5 demonstrates that in the five years before the patent library opening, there are no systematic differences in patenting between regions with Federal Depository Libraries that are about to receive a patent library and those regions with Federal Depository Libraries that do not obtain patent libraries. The number of patent applications that are ultimately granted increases in the years after patent library opening. In addition to not being present prior to the arrival of patent deposit libraries, the effects we find in (1) are also not evident in regions outside of the patent library’s commuting radius. Column (2) of Table 4 shows that the increase in patents is localized in the geographic region most proximate to the arriving patent library. For patents filed between 15 and 50 miles away we do not find any effect. In Appendix B.4 we report different distance bands and find no effect more than 15 miles away.<sup>23</sup>

---

<sup>22</sup>In Appendix B.5 we additionally drop each library in turn and find that the effect does not depend on any particular library in our main sample.

<sup>23</sup>In unreported regressions available upon request, we also use all libraries between 1975 and 1997 as the treatment group. This means that we also use patent libraries that are not FDLs and we add small FDLs as the control group. As a result, we have 56 rather than the baseline sample of 48 libraries in the treatment group. The effect is more than twice as high as in our main specification, but the coefficient becomes insignificant. In further unreported results, we have explored the data using count models, log specifications and different weighting schemes and find that the results do not depend on the particularities of the estimation method.

Table 4: Auxiliary Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					FDL Control group				Other Control	
					Add matching on		Distance Match		Groups	
	Base- line	Patents in 15-50 mi	Pseudo opening	Uni- versity	University Patents p.c.	University Patents p.c. Popu- lation	< 100 mi	5 clos- est	Syn- thetic	Within
Post	-0.2 (0.6)	-0.2 (0.4)	0.1 (0.5)	0.0 (0.7)	0.2 (0.5)	0.2 (0.5)	-0.8 (1.2)	-0.9 (1.3)	-0.5 (0.6)	-0.0 (0.1)
Pat Lib x Post	2.5** (1.2)	0.2 (0.5)	-0.3 (1.5)	3.2** (1.4)	2.3** (1.1)	2.3** (1.1)	3.2** (1.4)	3.7* (2.2)	1.5* (0.9)	1.7** (0.8)
Mean Dep.	14.5	3.2	14.0	13.4	11.9	12.4	16.5	15.5	17.4	13.0
R2 (within)	0.09	0.07	0.10	0.08	0.11	0.11	0.11	0.11	0.40	0.19
Obs.	4994	4994	4290	2948	2794	2640	3146	2629	1056	10740

**Note:** 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:

$$\frac{\#Patents_{it}}{Population_i} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within the same state as controls. In column (2) we use the number of patents between 15 and 50 miles as outcomes. In column (3) we assign a treatment indicator to the FDL closest to the patent library and drop all patent libraries from the sample. In column (4) we additionally use only FDLs that are also university libraries as controls if the patent library is also a university library. If the patent library is not a university library we only use FDLs that are not university libraries as controls. In column (5) we match on state, being a university and in addition we do a coarsened exact matching (CEM) with 5 bins on patent per capita in the year before the opening. In column (6) we add repeat the analysis of column (5) but also employ coarsened exact matching with 5 bins on population. Thus only FDLs in similar sized cities, with similar number of patent per capita and the same type of library are used as controls. In column (7) we do not match on state but take all FDLs within a 100 miles as controls. In column (8) we use the 5 closest FDLs as controls. In column (9) we construct a counterfactual by keeping the share of patents around patent libraries among U.S. patents constant to pre-opening levels. In column (10) we use not-yet opened patent libraries as control for opened patent libraries. 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 (assigned) patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

If a library had applied to become a patent depository library based on an expected increase in its local innovative capacity, the responsible librarians must have been able to pick exactly the right time and the right geographic location where patenting among local firms, including, in particular young firms, some of which were not yet founded, would increase. This foresight strikes us as less likely than the possibility that access to patent technical information supports the translation of ideas into patent applications.

### **No Effect of Pseudo Treatment of Closest Control Library**

Another prospect that threatens the interpretation that patent library opening induces an innovation response is the possibility that library opening may cause inventors to move geographically, but achieve no net effect on innovation. Were inventors to relocate to patent library regions, our results could reflect a simple change in the spatial distribution of patenting rather than an increase in innovation. If this were the case, the treatment assignment would then violate the stable unit treatment value assumption (SUTVA), as the opening up of the patent library would decrease innovation in control library regions. To test this possibility, we re-run our analysis assuming that the inventors most likely to relocate to the patent library regions are those in the geographically closest control regions. We therefore replicate (1) omitting patent libraries and, instead, assigning a fake treatment indicator to the Federal Depository Library closest to each dropped patent library. We report the result in column (3). The fact that the coefficient is neither statistically significant nor of a substantial magnitude suggests that there is no differential trend between closer control libraries and libraries that are further away. We conclude, therefore, that this type of interference is not a primary concern for this study.

### **Alternative Control Groups: Variants of FDL Matching and “Synthetic” Patent Libraries**

In columns (4) through (6) we explore the robustness of the results to the use of increasingly strict control groups. In column (4) we match on both state affiliation and on the status of being a university library. In column (5) we match on state and university and employ coarsened exact matching to ensure similarity in patenting per capita in the year before the opening, using five bins for patents per capita. In column (6) we use the same matching approach as in column (5) and also match on population within 15 miles of the library using five bins. In each case, the narrower control sample yields a reduction in the number of observations but a similarly-sized mean effect. In column (7) and (8) we do not condition on state but rely on a control group of FDLs defined by their distance to the patent library.

In column (7) we use the closest FDLs that are within 100 miles. In column (8) we use the five closest control libraries. In column (9), we construct the counterfactual by computing the “synthetic development” of patent library regions holding their share of a region among all U.S. patents constant to the pre-opening level.

### **Alternative Control Group: Not-yet-opened Patent Libraries**

Thus far, our analysis relies on the assumption that treatment and control regions differ only as a result of treated regions receiving patent libraries and, hence, the control regions enable us to estimate a counterfactual for the patenting that would have occurred in the absence of the patent libraries. We explore the robustness of our results to the relaxation of this assumption, using only treated libraries in the analysis in column (10). Specifically, we report the results of a model that identifies the impact of patent library opening using regions around not yet opened patent libraries as controls for patent libraries that were opened earlier (“Within sample”). In some sense, this is our strictest test since the identification here relies only on the timing of the treatment and not on any differences between libraries that receive patent collections and those that do not. Column (10) shows the effect on the number of patents per capita, our main measure of innovation. We again see a similar effect to our baseline specification. We explore the robustness of the other principal results to the use of this control group in Table 5. Column (1) repeats the last column of the previous table, showing the effect on patents per 100,000 population. Column (2) demonstrates that the effect on citation-weighted patents is smaller and not statistically significant, though it remains positive. When using the different measure of patent value by Kogan *et al.* (2017), we find positive and significant effects, in line with the main table. Together, these analyses suggest that, even in this more restrictive sample set, the results appear to reflect a set of economically meaningful innovations. Columns (4) and (5) show that the effect is larger for young than for old companies. Again, the effects are only statistically significant for young assignees and are substantially larger relative to their mean patenting rate. Columns (6) and (7) show that in the within-sample, the effect is positive and statistically significant for small companies, although it is insignificant (and even negative) for large companies. These results reinforce the importance of young and small companies in the boost in patenting after patent library opening.<sup>24</sup> Taken together, these additional analysis document the substantial robustness of the findings to the choice of control group. In Appendix B.2 we report average patenting per capita around patent library opening for these control groups.

---

<sup>24</sup>The persistence of our treatment effects over time throughout all results additionally suggests that salience is not the driving force behind our estimates.

Table 5: Using Future Patent Libraries as Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline			Age		Size	
	Patents	Citations	Dollars	Young	Old	Small	Large
Post	-0.0 (0.1)	-0.5 (2.4)	-4.2 (3.8)	-0.0 (0.0)	0.0 (0.1)	0.0 (0.1)	-0.0 (0.1)
Pat Lib x Post	1.7** (0.8)	13.4 (17.3)	42.3* (21.2)	1.1*** (0.4)	0.6 (0.6)	1.7*** (0.4)	-0.1 (0.6)
Mean Dep.	13.0	166.7	54.9	5.9	7.0	1.3	11.6
R2 (within)	0.19	0.20	0.23	0.19	0.09	0.33	0.12
Obs.	10740	10740	10740	10740	10740	10740	10740

**Note:** 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:

$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. We use not-yet opened patent libraries as control for opened patent libraries. In column (2) we weight each patent with its forward citations. In column (3) we use the patent values from Kogan *et al.* (2017) to weight each patent with its dollar value. We windorize the Dollar values at the 90th percentile to adjust for outliers. In columns (4) and (5) we split the dependent variable by young and old assignees. An assignee is young if it filed its first patent no more than three years before the opening of the library and old otherwise. In the following two columns we split the dependent variable by the size of assignee. An assignee is defined as large if it has more than 20 patents before the opening of the patent library. 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 (assigned) patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

## 5 Mechanism: Better Access to Patented Knowledge

Our prior analyses document that the opening of patent libraries induces an increase in local patenting. In this section, we explore whether better access to technical information disclosed in patents plays a key role driving these effects. We first recount the story of the development of Zithromax to illustrate that, if patents are informative, easier access to patents might improve the ability to build new technologies based on prior patented knowledge. Consistent with this disclosure mechanism, we find that the effect is driven by patents in the field of chemistry, in which patents provide extremely clear disclosure, and that inventors started to build on more geographically and technologically distant patents after a library opened. We then document that the impact of patent libraries on local innovation varies with the introduction of information technologies that affected the ease of prior art searches. Finally, one could be worried that patent libraries just substitute patent attorney activity. In Appendix B.6, we show that the number of local registered patent attorneys did not change in response to patent library openings.

### 5.1 An Illustrative Example: Patent Disclosure and the Development of Zithromax

Although academic analyses of patent libraries is scant, inventor accounts suggest that the disclosure of prior art via patent libraries facilitated inventive activity. Historical reports suggest that, in the early days of patent libraries, Thomas Edison made use of them to search for prior art (Sneed, 1998). Jack Kilby, the co-inventor of the integrated circuit, was expansive in his efforts to read patent documents issued by the U.S. government: “You read everything- that’s part of the job. You accumulate all this trivia, and you hope that someday maybe a millionth of it will be useful” (quoted in Stephan, 2012, p.226, from Reid, 1985).

The case of Zithromax provides an example of how this mechanism might work in practice (Idris, 2002; Li, 2009). Beginning in 1974, Pfizer had undertaken a program to develop a new macrolide, an antibiotic of the same type as erythromycin, but with greater antimicrobial effect. Despite significant investment, more than 2000 tested compounds and eight human trials, the firm did not make material progress. As a result, the firm was on the verge of closing down the program in 1980. While reviewing patent documents at the United States Patent and Trademark Office, Pfizer’s chemists stumbled upon a patent for a molecule with precisely the features they desired, which had been granted to the then-Yugoslavian firm, Pliva. Building directly upon the initial patent, two Pfizer chemists, G. Michael Bright and Dick Watrous, methylated an amine of Pliva’s drug, thus creating a slightly-modified version of the molecule, which Pfizer then patented. The time between the publication date of

Pliva’s patent and the application date of Pfizer’s patent was only six months. Subsequent to its patent filing, Pfizer reached a licensing agreement with Pliva and in 1991 received FDA approval to offer Azithromycin, for sale in the United States under the branded name, Zithromax. During the 1990s, it became one of the best selling branded antibiotics in the United States and worldwide, with total sales peaking at US\$2 billion in 2005.

## 5.2 Technology: Effect is Concentrated in Chemistry

If the effect in our principal analyses is, indeed, driven by improved access to patent prior art, we would expect the effect would be concentrated in technologies where patents are particularly informative. Evidence on this is presented by Gambardella *et al.* (2011), who report the results of surveys asking inventors how much time they saved by reading patents in various fields. They find that the average inventor claims to save twenty-five hours by reading patents in Chemistry, eight hours in Process Engineering, five hours in Instruments and Mechanical Engineering, three hours in Electrical Engineering, and eight hours in Other Fields. These results suggests that patents are most informative in chemistry. One reason may be that patents on chemical compounds display the specific molecular formations, thus fully disclosing the invention at hand. To illustrate this point, Appendix C.1 shows the patent on Acetyl Salicylic Acid, commonly known by its trade name, Aspirin, and displays the formula for the molecule. Due to the clarity of chemical disclosure and the clarity of the associated patent rights, chemistry is a field in which patents have been documented to be valuable and important for appropriability (Cohen *et al.*, 2000).

In Figure 6 we report the results of estimating Equation (1) using patents by technological fields as outcome variable.<sup>25</sup> To define technology fields we use a classification that aggregates IPC technology classes to larger sub fields (Schmoch, 2008). This is the same technology classification as in Gambardella *et al.* (2011). We find that the increase in patents is most pronounced in the field of Chemistry.<sup>26</sup> There is also smaller effect in Instruments. Consistent with the prospect that patent disclosure is a key mechanism driving the results, the effect is largest for inventors in fields where patents are most informative.<sup>27</sup>

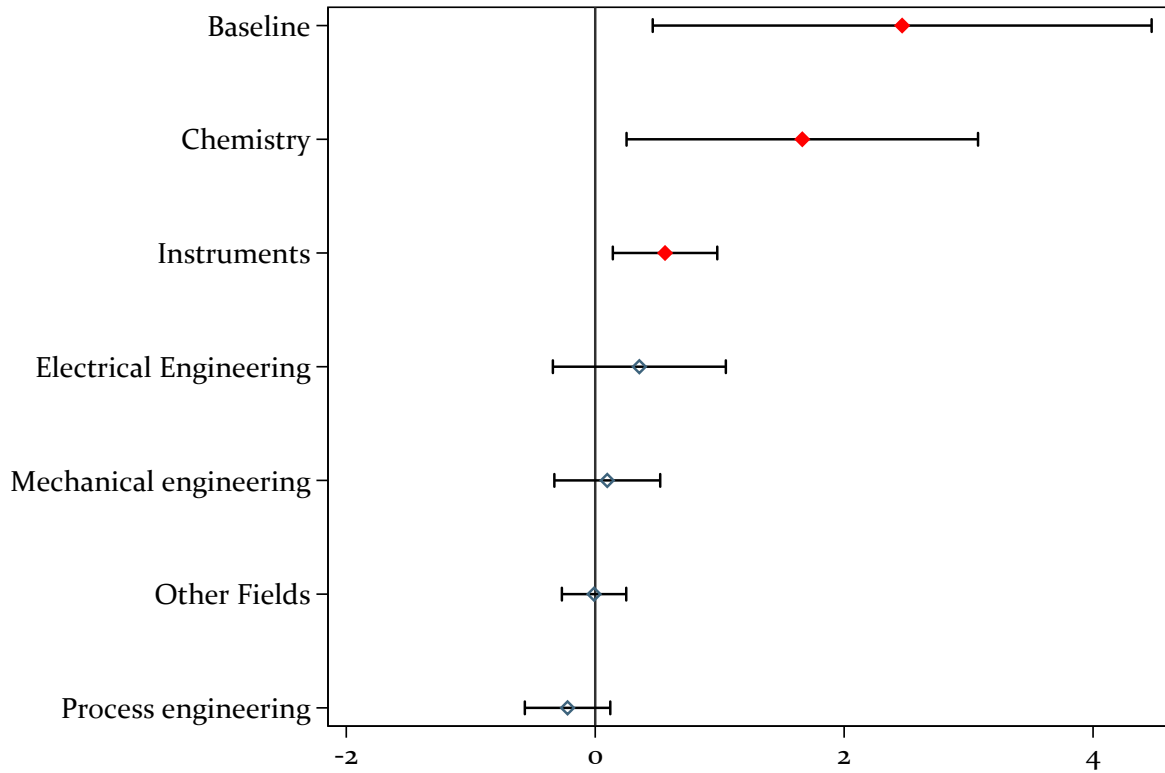
---

<sup>25</sup>Note that this specification splits the main dependent variable by field. Thus, the coefficients represent impacts on subsets of the number of patents per capita and therefore add up to the main effect.

<sup>26</sup>In Appendix C.2, we show that this effect only arises after the opening of the patent library.

<sup>27</sup>We replicate these results in Appendix C.3, using a different technology classification scheme that is more detailed and includes a larger number of fields. In unreported regressions, we also find that when using not-yet-opened patent libraries as control regions, the effects are similarly concentrated in chemistry.

Figure 6: Effect by Technology Category



**Note:** This figure 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:

$$\frac{\#Patents_{it\tau}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $\#Patents_{it\tau}$  is the number of patents within 15 miles of the library in year  $t$  and technological field  $\tau$ .  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. The technological fields follow the ISI-OST-INPI classification of 1995 as defined in Schmoch (2008). The range plots indicate the 90% confidence intervals for the coefficient that are plotted with a hollow diamond if the coefficient is not significantly different from zero or a full diamond if the coefficient is significantly different from zero. In Appendix C.3 we report the results for more detailed and alternative classifications of technological sub fields.



### 5.3 Structure of Patents: Citation Distance Increases, Patent Quality is Unchanged

If the arrival of patent libraries in a region truly induces innovation, such changes may be observable in changes in patent bibliometrics following patent library opening. For example, if these libraries extend the geographic reach of knowledge of distant patents, we would expect that this would make itself evident in an increase in the average distance to cited patents.

To investigate this possibility, we compare bibliometric features of patents associated with inventors in patent library regions with control patents of the same technology field and the same filing year but that were filed by inventors in Federal Depository Library regions. We again use the difference-in-differences specification in Equation (1), estimating now at the patent level, and asking how the nature of backward references and forward references change after library opening. We estimate each specification once for all patents of young companies (Panel A) and once for old companies (Panel B). We cluster standard errors at the patent library level.

Table 6 reports the results of models assessing the impact of library opening on the nature of patents in affected regions for young and old companies. Column (1) shows that the average number of backward citations increased for young firms. Induced patents may, thus, have profited more from prior art. We explore the geographical range of patent citation in column (2) by examining how library opening affects the mean geographic distance between citing and cited inventor.<sup>28</sup> Patents of young companies experience an increase in backward citation distance, consistent with the interpretation that library opening eased the difficulty that they previously faced in accessing prior art. The effect size corresponds to an increase of around 4% in distance across the set of patents in a region. There is an insignificant (negative) effect for old companies in Panel B. These results are consistent with what we would expect if patent access for previously-inhibited inventors was the driving mechanism behind the core findings, i.e., that patent libraries improve the access to distant and therefore less likely to be known patents. The results in columns (3) and (4) suggest that patents produced after patent library opening are also more original, i.e., cite more technologically-distant prior art.<sup>29</sup> We interpret these results as consistent with the prospect that the mechanism behind the post-library patent boost is the improved access to previously-distant and expensive-to-access prior art.

One other issue worth exploring is the possibility that library opening does not induce

---

<sup>28</sup>In Appendix C.4 we show the backward citation results for young firms by percentiles in the distance distribution.

<sup>29</sup>We define patent originality based on Hall *et al.* (2001).

Table 6: Impact of Patent Libraries on Structure of Patents

	(1)	(2)	(3)	(4)	(5)
Young companies					
	Backward citations	Mean backward distance	Originality	# Fields	Forward citations
Pat Lib x Post	0.2* (0.1)	40.8** (19.4)	1.8** (0.9)	3.7 (4.1)	1.3 (1.5)
Obs.	144446	144446	144446	144013	144446
Old companies					
	Backward citations	Mean backward distance	Originality	# Fields	Forward citations
Pat Lib x Post	-0.1 (0.2)	-18.6 (13.4)	1.0* (0.5)	-0.8 (1.6)	-0.3 (0.9)
Obs.	218177	218177	218177	217462	218177

**Note:** 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:

$$Outcome_{jt} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_j \cdot Post_t + Controls + \varepsilon_i$$

where  $Outcome_{jt}$  is the outcome for the patent  $j$  that is filed in year  $t$ .  $PatLib_j$  is an indicator if patent  $j$  is around a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. In column (1) we use the sum of backward citations. In column (2) we use the mean distance between the location of the inventor of the cited patent and the citing patent  $j$ . In column (3) we use originality of the patent as defined by Hall *et al.* (2001) and in column (4) we count the number of technical fields cited by the patent. In column 5 we use the sum of forward citations. The classification of technical fields follows Schmoch (2008). In column (1) we use a fixed effect for each combination of patent library, technology area and filing year as controls. In column (2) to (5) we use a fixed effect for each combination of patent library, technology area, filing year and number of backward citations as controls. In Panel A we use only companies with their first patent less than three years before the opening of the patent library. An old company is a company with a patent more than three years before the opening. Standard errors are clustered on the patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

innovation, but may simply cause a rush to submit any patentable invention. If library openings were to induce low quality patents, we would expect post-library patents to receive fewer forward citations than before. We investigate this question in column (5). The results evidence no decline in the number of forward citations, suggesting that induced patents are of similar quality (and value) to those produced before library opening.

#### 5.4 Over Time: Effect Decreases when Patents Become Available on the Internet

If, indeed, the disclosure of technical information were the relevant mechanism driving post-library patenting, we would expect that access to information technology would enhance the impact of patent libraries on local innovation. Prior to 1982, most documents in patent libraries were available either on microfilm or on paper but were not easily searchable. In 1982, the patent libraries introduced the computer-based database “CASSIS” to support prior art searches (Oliver, 2002, p.101). The search capabilities were “minimal in today’s time frame, but a quantum leap in 1982” (Sneed, 1998). Beginning with the patent database of the USPTO in the mid-1990s and the Espacenet database of the EPO in 1998, patent information became widely-available at low cost via the internet.<sup>30</sup> On December 14, 2006 Google Patents launched and initiated the era of ubiquitous access to easily searchable patent information. Considering this timeline, we hypothesize that the effect of opening patent libraries would be stronger beginning with the introduction of CASSIS in 1982 and that it would disappear around 1995, as a result of the release of the first USPTO patent database.

To investigate whether the effect of opening a patent library evolves over time, we estimate the Equation (1) separately for these different milestones of patent search capabilities. In Figure 7 we report the resulting treatment effects. For the period 1976 to 1981, the effect is indistinguishable from zero. During the period from 1982 to 1994, during which on-site databases were available but online databases are not, the effect is similar in magnitude to the baseline effect and statistically significant. The effect declines when online patent databases become available. These results are consistent with the expectations we outlined above, i.e., that the impact of libraries increased when patent depository libraries added searchable databases in the 1980s and then declined when patent information became available over the internet. It is important to note that, although the availability of patent information

---

<sup>30</sup>As late as 2004, however, experts said that both free databases “still tend to have primitive search engines and in some cases rather cumbersome mechanisms to download patents – deliberately so, on the part of the USPTO and EPO, who have said they do not wish to compete unfairly with commercial vendors,” implying that the library databases retained advantages relative to the free, internet-based versions (Lambert, 2004).

outside of PDL facilities changed in the late 1990s, patent libraries continued to operate in that period as they had in the early 1990s.<sup>31</sup>

## 6 Did Patent Libraries Affect Small Businesses?

If patent libraries improved the innovative productivity of local companies, we would expect to see this affect business statistics other than simply patenting. In particular, we would expect to see increased entry, survival, and growth of young and/or small enterprises. We investigate this question using U.S. Census data that report business statistics for Metropolitan Statistical Areas (MSA) between 1977 and 2005.<sup>32</sup> These data report MSA-level establishment entry, exit, job creation, and job destruction. We focus on firms with fewer than ten employees and on the subset of data that are not imputed for confidentiality reasons Miranda and Vilhuber (2014). We scale each variable by population (in 100,000). As we do not observe the exact location for each company, we compare MSAs with new patent libraries to MSAs with FDLs within the same state. While this approach results in a less precise comparison than our prior analyses, it nonetheless provides interesting insights into the potential effects of innovation as a result of improved local access to patent technical information.

To estimate the impact of library opening on the battery of outcomes available in the 1977-2005 MSA-level Census data, we use our standard differences-in-differences model. We plot the coefficient for the treatment effect for each outcome variable, along with its associated 90 percent confidence interval, in Figure 8. Panel A reports the results for establishments of firms with less than 10 employees. We find that library openings generate more establishment entries and job creation, in particular for new companies.<sup>33</sup> However, the effects for exit and job destruction (particularly by continuing firms) are also positive. As a result, net job creation has a positive mean coefficient, but is not statistically different from zero. In Panel B, we consider firm establishments with more than 10 employees. Consistent with the expectation that access to patent documents had a greater effect for smaller firms than for larger firms, we find no significant effect on entry or exit for these larger organizations. In Appendix C.7, we show selected regression results from these specifications.

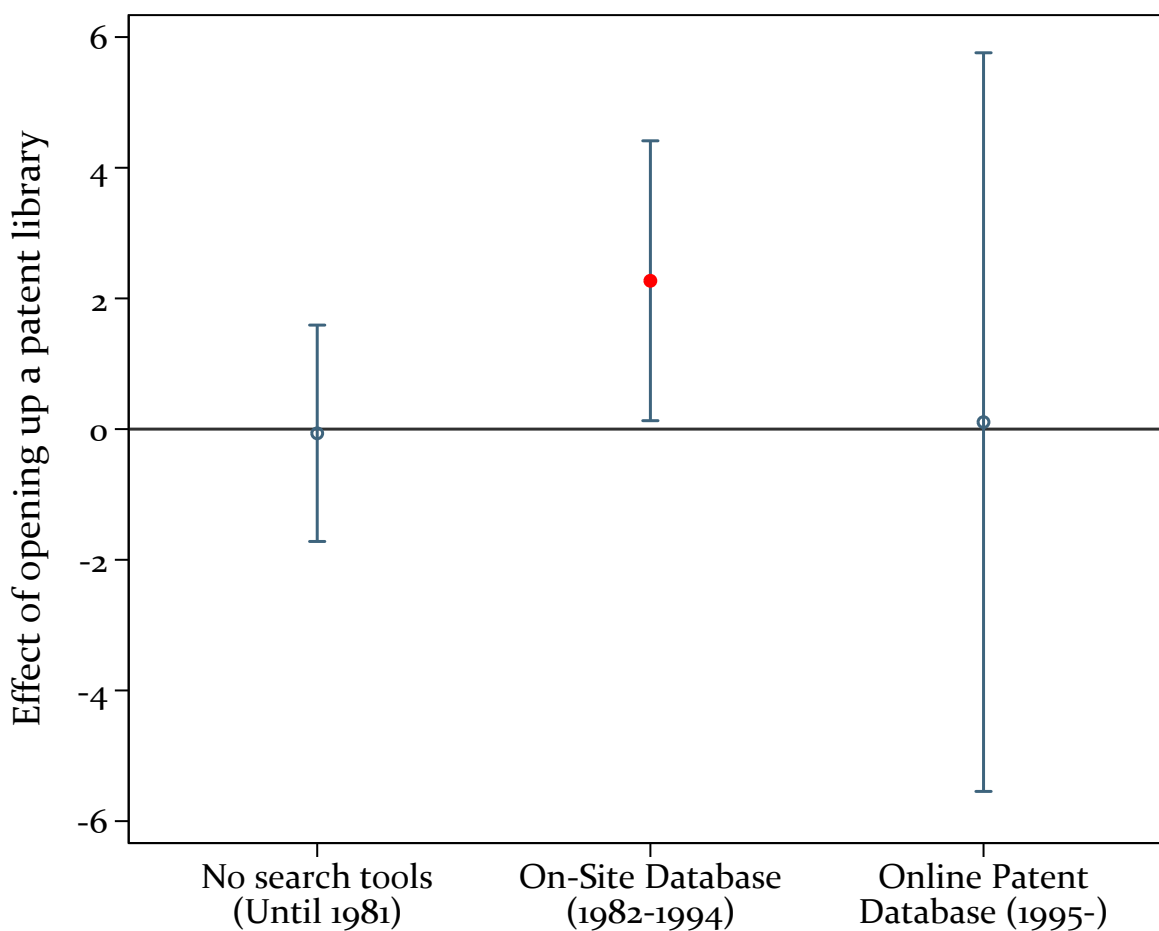
---

<sup>31</sup>In Appendix C.5, we additionally show that the long-run effects of opening patent libraries are substantial.

<sup>32</sup>The data are available via [https://www.census.gov/ces/dataproducts/bds/data\\_estab.html](https://www.census.gov/ces/dataproducts/bds/data_estab.html).

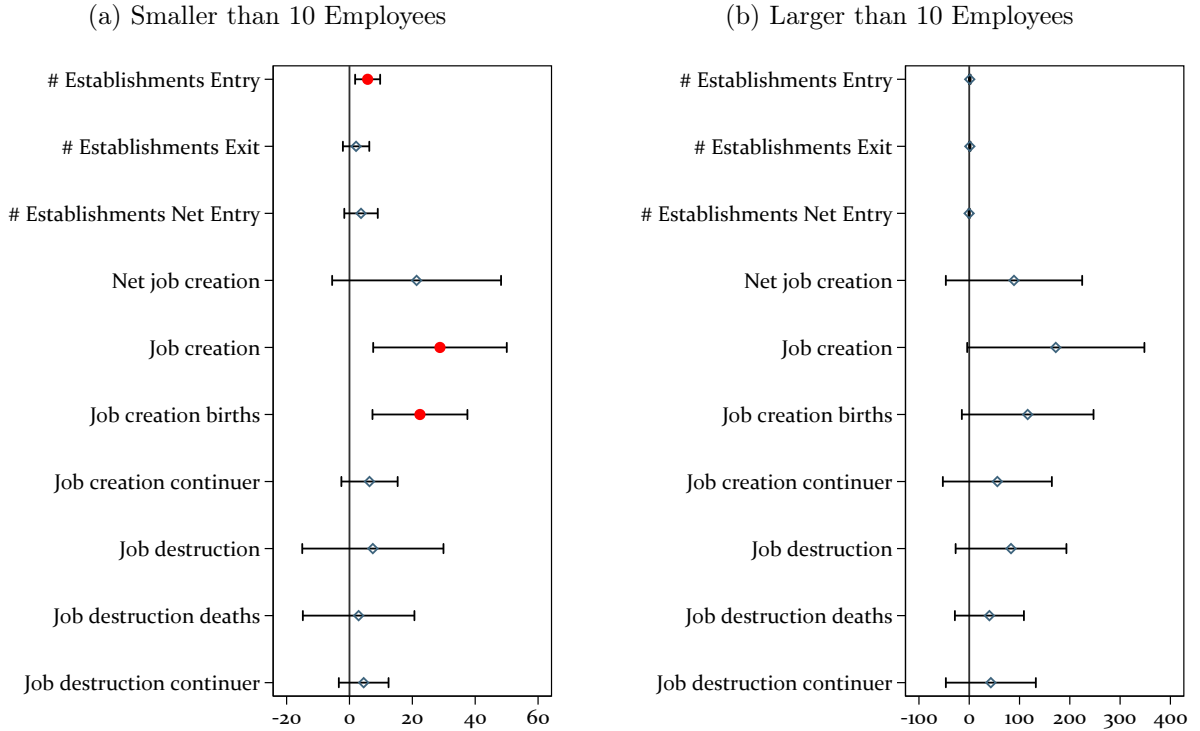
<sup>33</sup>To show the robustness of these estimates, in Section C.6 in the Appendix we repeat our leave-one-library-out analysis for our main result on establishment entry. The effects do not depend on any particular library. The results on all other significant measures are robust as well and are available on request.

Figure 7: Treatment Effect by Time Period in which Patent Library was Opened



**Note:** This figure shows the average treatment effects on the treated of opening up a patent library on the average number of patents relative to the average number of patents around matched federal depository libraries separately for the period before the introduction of on-site databases (until 1981), during the availability of these databases (1982-1994) and after the introduction of online patent databases (from 1995). For the time period after 1995, we include patent libraries that were not FDLs to have sufficient observations. We drop the library of Rochester NY because it has an extremely high patent per capita ratio as Rochester hosts Kodak, Xerox, and Bausch & Lomb and thus we cannot find a suitable control library.

Figure 8: Difference in Difference: Census Data



**Note:** This figure 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. In Subfigure a) the sample are all establishments smaller than 10 employees and in Subfigure b) all establishments larger than 10. The estimation equation is:

$$\frac{\#Outcome_{i,t}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $\#Outcome_{i,t}$  are the various outcome variables from the census in the MSA of the patent library.  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects.

Together, these results suggest that patent libraries contributed to small firm growth and the reallocation of workers and assets among firms. This is consistent with a process of creative destruction following increased innovative activity. In response to the opening of patent libraries, some entrants appear to have grown at the expense of incumbents. In these data, there seems to be a slight net positive effect for job creation in small companies and an increase in the number of small companies. Consistent with Furman and Stern (2011) and Biasi and Moser (2016), we conclude that patent deposit libraries act as knowledge-enhancing institutions whose arrival constitutes an improvement in, and potentially a catalyst for, the local innovation environment.

## 7 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. Courts and scholars argue that disclosure constitutes a significant benefit of the patent system, as it helps inventors to avoid duplication and gives them new ideas to recombine with their own. However, there is very little evidence whether or not disclosure resulting from the patent system affects innovation. By leveraging geographic variation in access to patent prior art during a key period of expansion in the USPTO Patent Depository Library Program, we are able to shed light on this issue. Specifically, we document that the opening of patent libraries from 1975 to 1997 increased innovation in the regions in which those libraries opened. Consistent with the expectations of Machlup (1958) and Scotchmer and Green (1990), we find that the enablement effect resulting from the disclosure of knowledge contained in patents is quantitatively important for subsequent innovation and, further, that this effect is focused in those fields in which patent disclosure appears to provide valuable information to future innovators.

In addition to providing evidence regarding a key question in the economics of intellectual property, our study contributes to the literature on research enhancing institutions. While economists generally agree that institutions that lower the costs of access to useful knowledge may support innovation (Mokyr, 2002), empirical research has provided few examples (Furman and Stern, 2011; Biasi and Moser, 2016; Waldinger, 2016). In this work, we document the value of patent libraries as knowledge hubs whose operation contributes to follow-on innovation, enhances local innovation systems, and facilitates the democratization of innovation by supporting innovation among relatively younger and smaller enterprises. In addition to their direct benefits, these institutions create opportunities for resource-constrained firms to act in a strategic manner to benefit disproportionately from public investments.

## References

- ARORA, A., CECCAGNOLI, M. and COHEN, W. M. (2008). R&d and the patent premium. *International Journal of Industrial Organization*, **26** (5), 1153–1179.
- BALL, G. (2012). *No more laughing at the bad boy: A technological adventure between Silicon Valley and the Alps*. Haymon.
- BALSMEIER, B., LI, G.-C., CHESEBRO, T., ZANG, G., FIERRO, G., JOHNSON, K., KAULAGI, A., LUECK, S., O'REAGAN, D., YEH, B. *et al.* (2017). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy*.
- BARUFFALDI, S. and SIMETH, M. (2018). Patents and knowledge diffusion: The effect of early disclosure.
- BESSEN, J. and MEURER, M. J. (2008). *Patent failure: How judges, bureaucrats, and lawyers put innovators at risk*. Princeton University Press.
- BIASI, B. and MOSER, P. (2016). Effects of copyrights on science: Evidence from the wwii book republication program. *Available at SSRN 2542879*.
- BROWN, W. H. and ARSHEM, J. A. (1993). *Survey of patent and trademark depository libraries 1991/92*. U.S. Patent and Trademark Office, Washington D.C.
- CHIEN, C. V. (2016). Contextualizing patent disclosure. *Vanderbilt Law Review*, **69** (6), 1849–1890.
- COHEN, W. M., GOTO, A., NAGATA, A., NELSON, R. R. and WALSH, J. P. (2002). R&d spillovers, patents and the incentives to innovate in Japan and the United States. *Research Policy*, **31** (8), 1349–1367.
- , NELSON, R. R. and WALSH, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why u.s. manufacturing firms patent (or not). *NBER Working Paper 7552*.
- FROMER, J. C. (2008). Patent disclosure. *Iowa L. Rev.*, **94**, 539.
- FURMAN, J. and STERN, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, **101**, 1933–1963.



- GAMBARDELLA, A., HARHOFF, D. and NAGAOKA, S. (2011). The social value of patent disclosure, mimeo.
- GRAHAM, S. and HEGDE, D. (2015). Disclosing patents' secrets. *Science*, **347** (6219), 236–237.
- GROSS, D. P. (2018). Suspended benefits: Evidence on the incentive and disclosure functions of patents from the u.s. patent secrecy program in world war ii. *Working Paper*.
- HALL, B. H. and HARHOFF, D. (2012). Recent research on the economics of patents. *Annual Review of Economics*, **4**, 541–565.
- , JAFFE, A. B. and TRAJTENBERG, M. (2001). The nber patent citation data file: Lessons, insights and methodological tools. *NBER Working Paper No. 8498*.
- HEGDE, D. and LUO, H. (2017). Patent publication and the market for ideas. *Management Science*, **64** (2), 652–672.
- IACUS, S. M., KING, G. and PORRO, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, **20** (1), 1–24.
- IDRIS, K. (2002). *Intellectual property: a power tool for economic growth*, vol. 888. WIPO.
- 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** (1-2), 183–201.
- KHAN, B. Z. (2005). *Democratizing Innovation*. Cambridge University Press: Cambridge UK.
- KOGAN, L., PAPANIKOLAOU, D., SERU, A. and STOFFMAN, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, **132** (2), 665–712.
- LAMBERT, N. (2004). Internet patent information in the 21st century: A comparison of delphion, micropatent, and qpat. In *Proceedings of the 2004 International Chemical Information Conference*.
- LAMOREAUX, N. R. and SOKOLOFF, K. L. (1999). *Inventors, firms, and the market for technology in the late nineteenth and early twentieth centuries*, University of Chicago Press, chap. Learning by Doing in Markets, Firms and Countries, pp. 19 – 60.

- LANDES, W. M. and POSNER, R. A. (2003). Indefinitely renewable copyright. *The University of Chicago Law Review*, **70** (2), 471–518.
- LEE, W. F. and COGSWELL III, L. P. (2004). Understanding and addressing the unfair dilemma created by the doctrine of willful patent infringement. *Houston Law Review*, **41**, 393.
- LEMLEY, M. A. (2012). The myth of the sole inventor. *Michigan Law Review*, **110** (5), 709–760.
- LEVIN, R. C., KLEVORICK, A. K., NELSON, R. R., WINTER, S. G., GILBERT, R. and GRILICHES, Z. (1987). Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity*, **1987** (3), 783–831.
- LI, J. J. (2009). *Triumph of the heart: the story of statins*. Oxford University Press.
- MACHLUP, F. (1958). *An economic review of the patent system*. 15, US Government Printing Office.
- and PENROSE, E. (1950). The patent controversy in the nineteenth century. *The Journal of Economic History*, **10** (1), 1–29.
- MANSFIELD, E. (1986). Patents and innovation: An empirical study. *Management Science*, **32**, 173–181.
- MIRANDA, J. and VILHUBER, L. (2014). Using partially synthetic data to replace suppression in the business dynamics statistics: early results. pp. 232–242.
- MOKYR, J. (2002). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.
- MORRISON, G., RICCABONI, M. and PAMMOLLI, F. (2017). Disambiguation of patent inventors and assignees using high-resolution geolocation data. *Scientific Data*, **4**.
- MOSER, P. (2011). Do patents weaken the localization of innovations? evidence from world’s fairs. *The Journal of Economic History*, **71** (2), 43–74.
- (2013). Patents and innovation: Evidence from economic history. *Journal of Economic Perspectives*, **27** (1), 23–44.
- OLIVER, J. (ed.) (2002). *Celebrating 25 years of libraries and USPTO cooperation: 1977-2002*. United States Patent & Trademark Office.

- OUELLETTE, L. L. (2012). Do patents disclose useful information? *Harvard Journal of Law & Technology*, **25** (2), 546–602.
- (2017). Who reads patents? *Nature Biotechnology*, **35** (5), 421–424.
- PATENT, U. S. and OFFICE, T. (1999). *Survey of Patent and Trademark Depository Libraries: 1997 Customer Satisfaction Survey*. No. Bd. 21 in *Survey of Patent and Trademark Depository Libraries: 1997 Customer Satisfaction Survey*, U.S. Patent and Trademark Office.
- REID, T. (1985). *The chip: How two Americans invented the microchip and launched a revolution*. Random House.
- ROIN, B. (2005). The disclosure function of the patent system (or lack thereof). *Harvard Law Review*, **118**, 2007–2028.
- ROMER, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, **98** (5, Part 2), S71–S102.
- SCHMOCH, U. (2008). Concept of a technology classification for country comparisons. *Final report to the world intellectual property organisation (wipo)*, WIPO.
- SCOTCHMER, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives*, **5** (1), 29–41.
- and GREEN, J. (1990). Novelty and disclosure in patent law. *RAND Journal of Economics*, **21** (1), 131–146.
- SNEED, M. C. (1998). 125 years of patent information to the people: the us patent and trademark depository library program.
- (2000). Fully disclosed yet merely descriptive: Intricacies of training the patent and trademark information professional. *Journal of Library Administration*, **29** (1), 59–78.
- STEPHAN, P. E. (2012). *How economics shapes science*. Harvard University Press.
- WALDINGER, F. (2016). Bombs, brains, and science: The role of human and physical capital for the production of scientific knowledge. *The Review of Economics and Statistics*, **98** (5), 811–831.
- WILLIAMS, H. L. (2017). How do patents affect research investments? *NBER Working Paper No. 23088*.

## A Appendix to Sections 2 & 3

### A.1 List of All Opened Patent Libraries

Table A-1 and Table A-2 show a list of all Patent Depository Libraries in our data, following Jenda (2005).

### A.2 Dataset Construction

We process the patent data and the data on libraries in the following steps to arrive at our final dataset.

#### Patent Data

1. We use patent data from the PATSTAT Database of the EPO that contains the universe of U.S. patents.
2. We delete all patents that pertain to foreign inventors.
3. We geolocate all patents using the data of Balsmeier *et al.* (2017) and Morrison *et al.* (2017).
4. We account for patents with inventors in multiple cities by using city-weighted patents.
5. To calculate citation distance, we assign the address of the first inventor on the citing or cited patent to the entire patent. When there is no primary inventor, we keep the first one in the list. We use only citations that are within the U.S.
6. We use population data from U.S. Census at the level of the incorporated city and compute yearly patent and citation rates per capita in circles around all library locations.

#### Library Data

1. Data on patent libraries (see tables A-1 and A-2) are from Jenda (2005) and the complete list of Federal Depository Libraries is from the online Federal Depository Library Directory.

Table A-1: List of all Patent Depository Libraries

City, State	Name of Library	Opening 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, 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
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

Table A-2: List of all patent libraries (continued)

City, State	Name of Library	Opening Year
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

2. We drop the Federal Depository Libraries outside the continental United States, including Pago Pago AS; Mangilao GU; Saint Thomas VI; Kolonia, Pohnpei FM; and Saint Croix VI. We obtain the library location information based on their city and state.
3. We geolocate patent libraries and Federal Depository Libraries using patent data, as all patent libraries are in places with at least one patent between 1975 to 2005. We match all Federal Depository Libraries within 250 miles to a patent library. If a Federal Depository Library can be assigned to multiple patent libraries, we match it to the geographically closest patent library.
4. We drop all patent libraries that are not also Federal Depository Libraries. To obtain a better match of treatment and control library we delete all small federal depository libraries because patent depository libraries are usually either medium sized or large federal depository libraries. Of the patent libraries that were opened in our sample period that are also FDLs, 96% are considered medium sized or large, and only three patent libraries are considered small.

In a last step we cross all inventor locations with our library data to obtain pair-wise combinations of locations between inventors and patent libraries. For each inventor location, we thus have a closest library. Using this, we can assign a closest patent library to each inventor-patent combination.

## **B Appendix to Sections 3 & 4**

### **B.1 Summary Statistics without Outlier Regions**

In Table B-1 we show summary statistics of the sample after deleting outlier control regions that report zero patenting in at least one year. While the mean differences do not affect our assumptions in the difference-in-differences setup, deleting these regions improves the balancing. Our results are identical when excluding these outlier regions from our regressions.

### **B.2 Compare Averages**

In the main part of the paper we employ three fundamentally different approaches to constructing the control group for the patent libraries. In Figure B-1a we report the raw difference in the average number of patents per 100,000 persons around treatment and the three control groups. As a first control group, we use Federal Depository Libraries within

Table B-1: Summary Statistics in the Year Before Opening

<i>Main sample</i>				
	Patent Libraries	Control Libraries	Diff	P-Value
Population in 100k	7.60	8.26	0.66	0.78
Uni Library	0.67	0.70	0.03	0.71
# Patents	128.29	90.72	-37.56	0.18
# Patents/100k	15.68	12.56	-3.12	0.23
Citation-weighted patents	226.76	197.06	-29.70	0.54
Dollar-weighted patents	83.80	63.54	-20.26	0.36
# Pat. small firms/100k	7.25	6.22	-1.03	0.35
# Pat. big firms/100k	8.43	6.34	-2.08	0.30
# Pat. young firms/100k	5.45	4.48	-0.96	0.30
# Patents old firms/100k	10.23	8.08	-2.15	0.31
Number of libraries	48	342		
<i>Patents by field</i>				
	Patent Libraries	Control Libraries	Diff	P-Value
Electrical Engineering	2.23	2.19	-0.04	0.95
Instruments	2.27	1.97	-0.30	0.46
Chemistry	4.02	2.19	-1.82	0.15
Process Engineering	2.02	2.33	0.31	0.49
Mechanical Engineering	2.98	2.09	-0.89	0.28
Other Fields	2.14	1.77	-0.37	0.38

**Note:** This table shows the averages of the data for patent libraries and control libraries without outlier regions that report zero patenting in some years. 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 three years before the opening of the patent library, otherwise it is old. A firm is defined as small if it has no more than 20 patents before the opening of the patent library, otherwise it is large. The p-values result from a t-test with unequal variances.



250 miles around the patent library . As a second control group, we construct a “synthetic” doppelgaenger for each patent library. We do this by calculating how many patents would be around a patent library in the years after the opening if the average share of patents of the region among all patents in the U.S. would have remained constant. The third control group involves patent libraries that are opened at a future date (“within”). In this Figure, we use all future patent libraries as controls of the library to keep the sample constant. In the body of the paper and in Figure B-2, we drop patent libraries from the sample as soon as they are opened.

Before patent library opening, the number of patents around treatment and control are stable. After the opening of a patent library, there are more patents filed around the patent libraries. There is also an upward trend among all potential control groups. To better see the relative increase, we subtract from each series in Figure B-1b its value in the year before the opening of the library to account for different levels of patenting. As in Figure 4 in the main text, patent libraries increase the local patenting rate relative to all specifications of the control group.

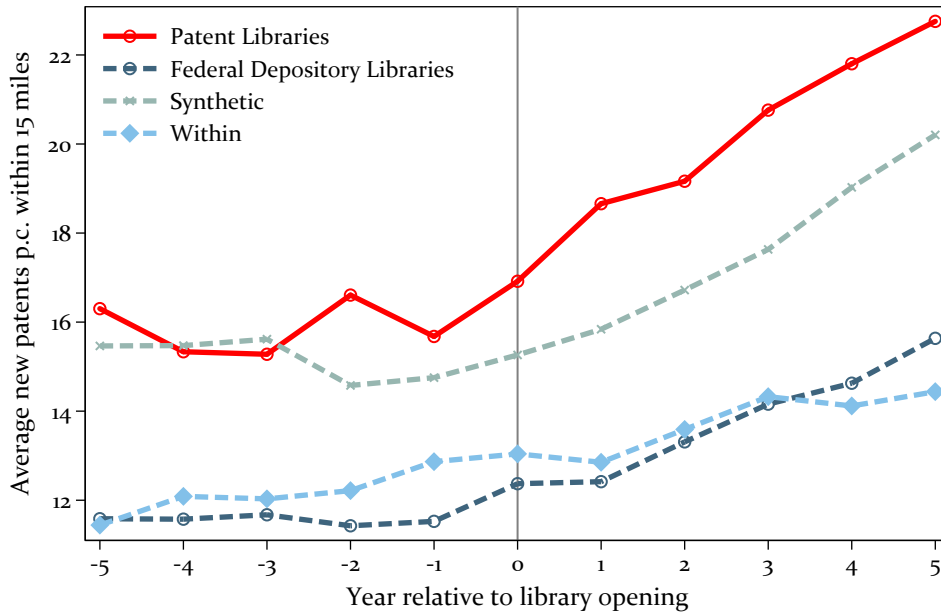
However, note that in the previous figure, the control group in the within-sample contains treated observations, namely patent libraries that open within five years of the opening of the patent library in the treatment group. We include these so that we do not have to drop patent libraries from the treatment group and can thus show an average that is comparable to the other control groups. In Figure B-2, we adjust this and again use future patent libraries as controls but drop them as soon as they are opened. If no control library is left we accordingly drop the patent library from the treatment sample. This is the within-sample used in our regressions in Table 5. Figure B-2a shows absolute values while Figure B-2b shows the the patent number normalized in the year before opening. Again, both lines diverge substantially after patent library opening, while being comparable before. Because both the number of treatment and control observations is diminished in this sample, we prefer to use Federal Depository Libraries as our main control group.

### **B.3 Further Sample Splits and Robustness**

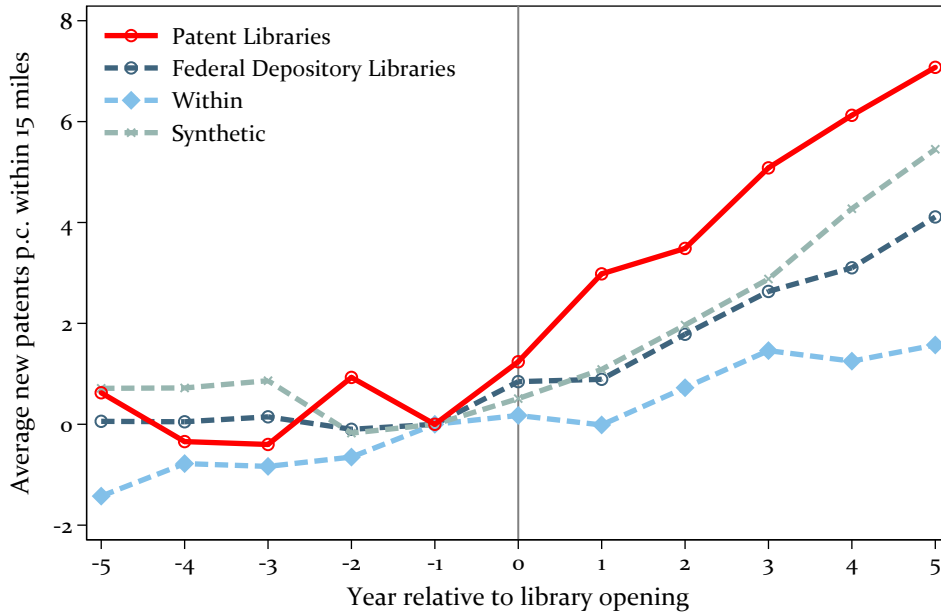
In Figure B-3 we report further results for our main specification. We first split the dependent variable by the type of assignee and find that the effect is driven by patents assigned to companies. To a smaller degree, the effect is also present for patents assigned to universities. In the last two lines, we split the sample in historically high and low patent regions. The effect is statistically significant only in historically low patenting regions.

Figure B-1: Compare Averages

(a) Raw



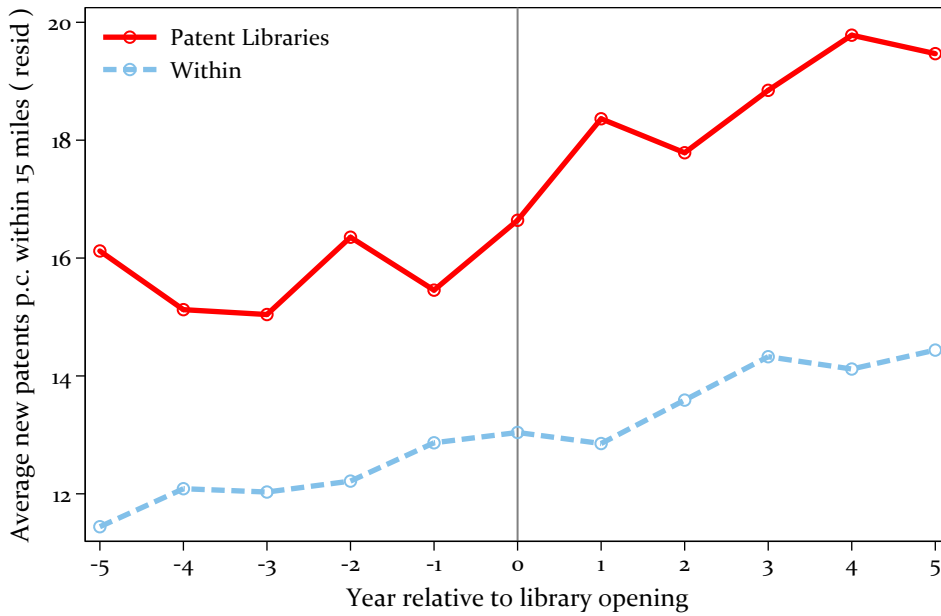
(b) Normalized in the Opening Year



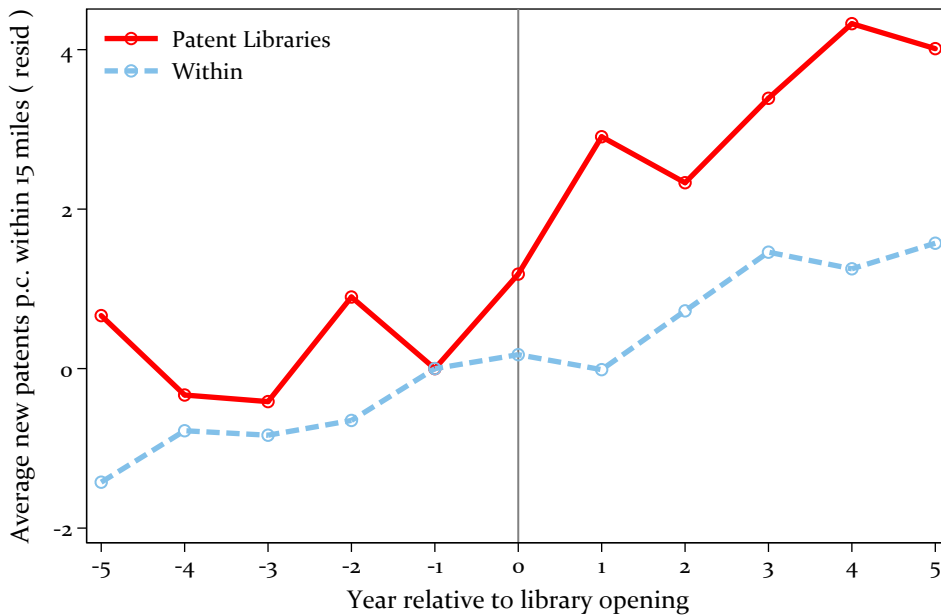
**Note:** This figure plots the average number of patents within 15 miles of the patent library (red solid line), around Federal Depository Libraries (dark blue dashed line), around synthetic patent library regions (green dashed line), and around patent libraries that are opened later (light blue dashed line) in the five years before and after the opening of the library. Figure B-1a shows the raw average and in Figure B-1b we normalize the average relative to its value in the year before the opening.

Figure B-2: Compare Averages: Within Sample

(a) Levels

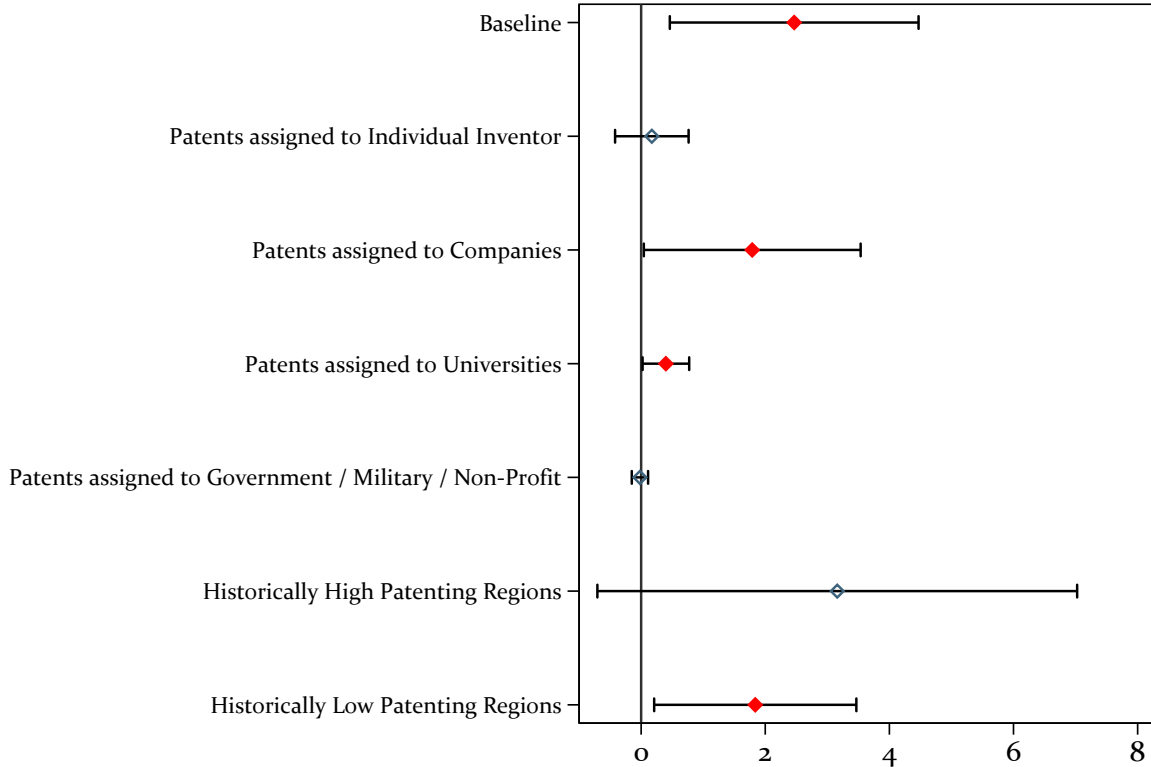


(b) Normalized in the Year before Opening



**Note:** This figure plots the average number of patents within 15 miles of the patent library (red solid line) and around matched not yet opened patent libraries (blue dashed line) in the five years before and after the opening of the library. Figure B-2a shows the raw average and in Figure B-2b we normalize the average relative to its value in the year before the opening.

Figure B-3: Further Main Results



**Note:** This figure 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:

$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In the first line we report the point estimate for  $\beta_2$  along with 90% confidence intervals. The confidence intervals are based on standard errors that are clustered on the patent library level. In lines (2) to (5) we split the dependent variable by the type of assignee. We show results separately for independent inventors, patents assigned to companies, patents assigned to universities, and patents assigned to the government, military or non-profits. In lines (6) and (7) we split the sample by an indicator if the region of the patent library has historically many or historically few patents. We define a region as having many patents if the average yearly number of patents per capita is above the median.

In Table B-2, we provide additional robustness tests for our main specification. In the first column we show our baseline estimate. In the second column, we repeat the estimation without using the weights of the CEM algorithm. In the third column, we include patent library-specific trends in the regression. In the fourth column, as the dependent variable we only use the per capita count of patents whose inventors are within 15 miles of each other or patents from solo inventors. In the final two columns, we show the robustness of our results when using larger circles around patent libraries to compute patents per capita with 25 and

50 miles, respectively. Finally, in the last two columns we use the number and the log number of patents in the 15 miles around the library as the respective dependent variable, controlling for (now time-varying) population. Our results are robust to all of these robustness tests.

## **B.4 Alternative Distances**

In addition to not being present prior to patent deposit library opening, the effects that we find in (1) are not evident in regions outside of the patent library's commuting radius. Column (2) of Table 4 and Figure B-4 show that the increase in patents is localized in a geographic region most proximate to the arriving patent library. For patents filed by inventors whose addresses are further than 15 miles from opened patent libraries, the impact of library opening is neither of substantial economic magnitude nor does it have a statistically significant effect. In this analysis, we consider the outcome variable to be the number of patents in a variety of distance bands around the treatment and the control libraries. This result implies that the number of patents only increases around the patent library but not in the wider area. Further, the finding increases our confidence that regions are not receiving patent libraries in anticipation of increasing innovation potential. If a region was chosen to get a patent library based on an expected increase in its innovative capacity, the government must have been able pick exactly the right spot where patenting will increase.

## **B.5 Leave-one-out Estimation: The Impact of Individual Patent Depository Libraries**

In our final set of analyses of the robustness of the results to alternative samples, we explore the role of individual library regions. In Figure B-5 we run our main analysis, dropping library regions one by one. With the exception of the library in Ann Arbor MI, we find that the coefficient indicating the post-patent library effect does not change. Dropping Ann Arbor reduces the coefficient from 2.5 to around 1.7, while making the estimate more precise though still within the initial confidence interval. As we described above, our main sample excludes the patent libraries of Burlington VT. This region has an extremely high patent per capita ratio because Burlington VT was the home of IBM's major research facility. This constitutes a substantial innovation outlier in its local area and, indeed, in the entire dataset. As a result, we could not identify a control region within 250 miles and within the same state that achieved even remotely similar levels of per capita patenting. When we add the library to our main analysis, we find a post library opening effect size greater than that in our preferred specification, but also that the additional noise renders the coefficient indistinguishable from zero.

Table B-2: Further Robustness Tests

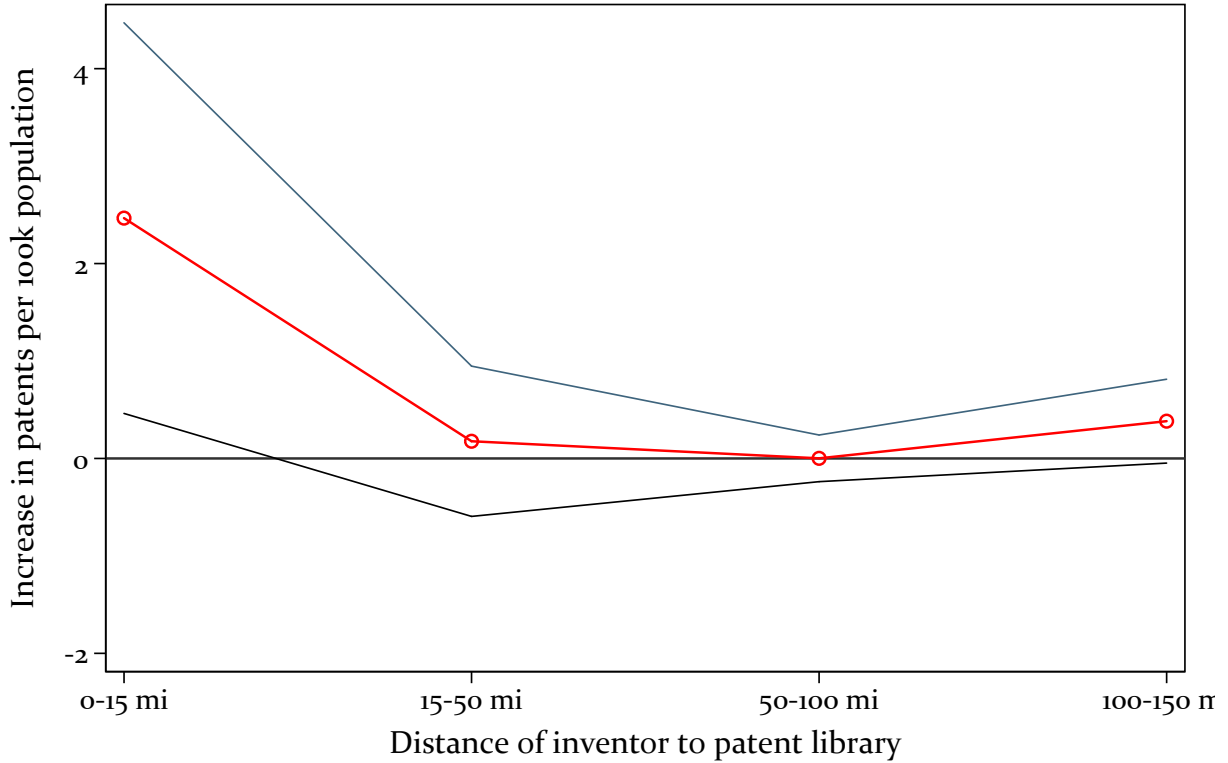
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline			Inventors who are..			# Patents	
	Baseline	W/o CEM Weights	PTDL spec. Trends	Co-located	25m around library	50m	# + Pop.	Ln(#+1) + Pop.
Post	-0.2 (0.6)	-0.1 (0.5)		-0.2 (0.5)	-0.4 (0.4)	-0.3 (0.2)	-1.6 (3.3)	0.0 (0.0)
Pat Lib x Post	2.5** (1.2)	2.4* (1.3)	2.5** (1.2)	2.1* (1.2)	1.8** (0.8)	1.1** (0.6)	16.8* (9.0)	0.1** (0.0)
Mean Dep.	14.5	14.5	14.5	11.3	11.2	6.2	94.3	3.3
R2 (within)	0.09	0.11	0.26	0.06	0.13	0.19	0.10	0.40
Obs.	4994	4994	4994	4994	4994	4994	4994	4994

**Note:** 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 baseline estimation equation is:

$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use patents within 15m of the patent library per 100,000 inhabitants as the dependent variable. In column (2) we repeat our baseline estimation without using the weights suggested by Iacus *et al.* (2012). In column (3) we include (assigned) patent library specific trends. In column (4) we only use patents by solo or co-located inventors within 15 miles from each other as the dependent variable. In the following two columns we use circles of 25 and 50 miles around the patent library to compute the dependent variable. In column (7), we use the number of patents in the 15 miles around the libraries as the dependent variable and control for (time-varying) population. In the final column, we repeat this estimation using the log number of patents as the dependent variable. In all regressions but column (2), 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 (assigned) patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure B-4: Effect of Patent Libraries by Distance

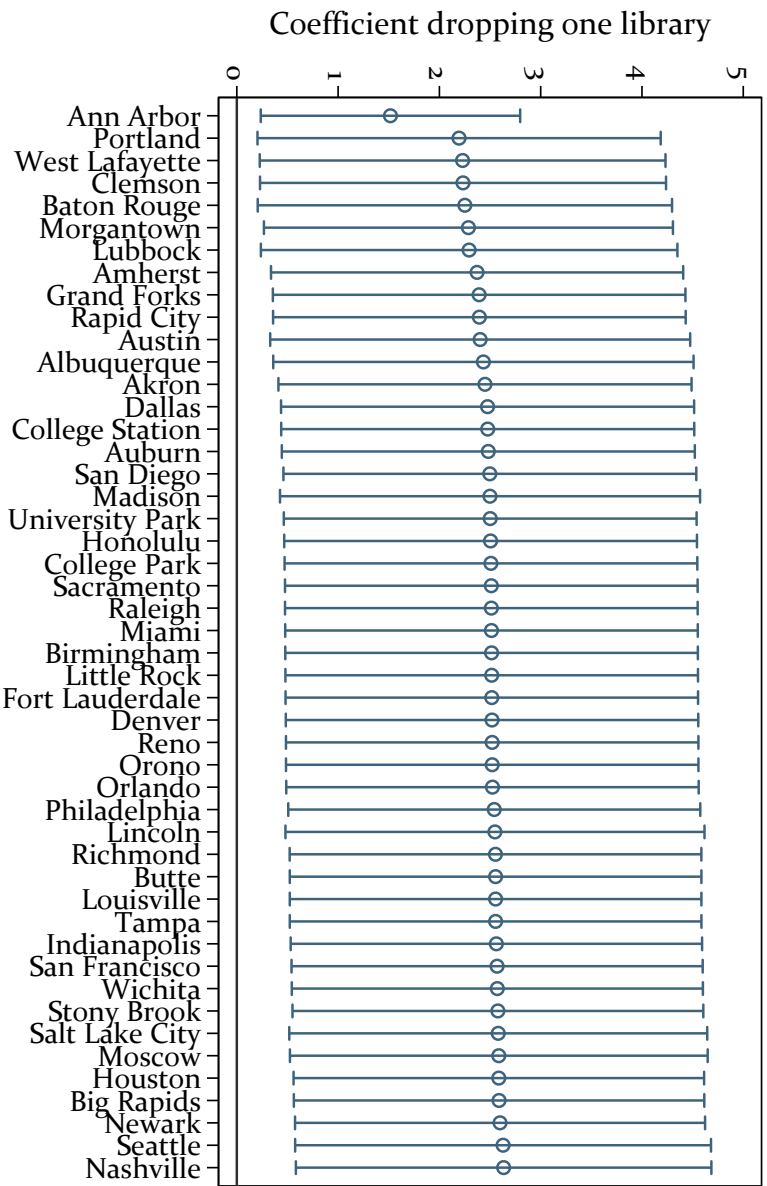


**Note:** This figure shows the coefficient  $\beta_2$  from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is:

$$\frac{\#Patents_{itd}}{Population_{id}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $\#Patents_{it}$  is the number of patents in distance  $d$  of the library in year  $t$ .  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. For each plotted coefficient we use the distance band reported on the horizontal axis. We report 90% confidence intervals for the coefficient. The confidence intervals are based on standard errors that are clustered on the patent library level.

Figure B-5: Stability: Leave-one-library-out Estimation



**Note:** This figure shows the coefficient  $\beta_2$  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:

$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \epsilon_i$$

where  $\#Patents_{it}$  is the number of patents within 15 miles of the library in year  $t$ .  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. For each plotted coefficient we leave out the patent library on the horizontal axis. The range plots indicate the 90% confidence intervals for the coefficient.



## B.6 Patent Attorney Results

In Figure B-6 we use data from the historical rosters of registered patent attorneys at the USPTO to provide evidence on the impact of patent libraries on the local number of active patent attorneys. In line with our identification assumption, the number of patent attorneys in treatment and control group is similar prior the opening of libraries. In addition, there is no clear effect of the opening of patent libraries on the number of registered attorneys at the USPTO. These findings suggest that an influx of patent attorneys is unlikely to induce patent opening and is also unlikely to account for the boost in patenting that follows library arrival.

We also document the robustness of the paper's core results to controlling for the number of local patent attorneys in Table B-3. The first column replicates our baseline estimates. The second column does so for the subsample of library-year observations where patent attorney data is available. The third column shows that for this subsample, controlling for the number of patent attorneys per capita does not affect our estimates. If anything, this increases the estimated impact of opening a library. In line with what we would expect, the number of patent attorneys per capita in a region positively predicts local patenting. In the fourth column, we analyze the impact of opening a patent library on the number of patent attorneys per capita. In line with Figure B-6, patent libraries do not seem to affect the number of active patent attorneys.

## C Appendix to Sections 5

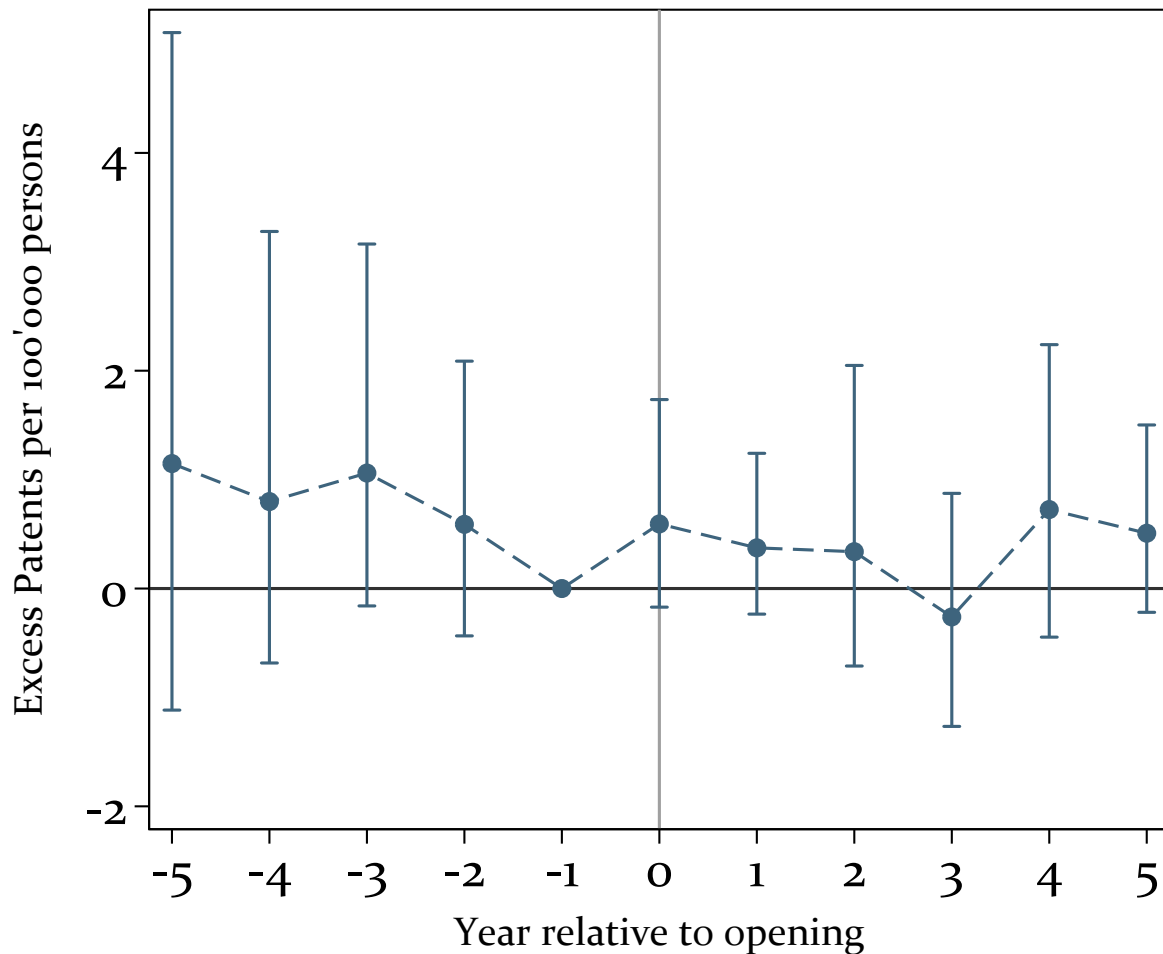
### C.1 Example for Chemical Patents: Aspirin

Figure C-1 shows the patent for Acetyl Salicylic Acid, commonly known by its trade name Aspirin.

### C.2 Time-Varying Treatment Effect for Chemistry

To assess whether our effects in chemistry are already present in the pre-period, we repeat our estimation of non-parametric treatment effects in Figure C-2, using patents in chemistry per 100,000 within 15 miles of a library as the dependent variable. As in our main estimates, the effect only arises after actual patent library opening.

Figure B-6: Impact on Number of Patent Attorneys p.c.



**Note:** This figure shows the yearly average treatment effects on the treated of opening up a patent library on the average number of patent attorneys within 15 miles of patent libraries relative to the average number of patent attorneys around matched federal depository libraries. The 90% confidence intervals (in blue) are based on bootstrapped standard errors. We use the weights of Iacus *et al.* (2012) to arrive at the average treatment effect on the treated. Data on patent attorneys comes from the historical rosters of registered patent attorneys from the USPTO.

Table B-3: Patent Attorney Results

	Dependent Variable			
	(1)	(2)	(3)	(4)
	Patents p.c.			Attorneys p.c.
Post	-0.2 (0.6)	0.6 (0.8)	1.1 (0.9)	-0.2** (0.1)
Pat Lib x Post	2.5** (1.2)	2.3* (1.1)	2.7** (1.0)	-0.1 (0.2)
Patent Attorneys p.c.			2.7** (1.3)	
Mean Dep.	14.5	15.3	15.3	2.0
R2 (within)	0.09	0.11	0.23	0.15
Obs.	4994	2848	2848	2848

**Note:** 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:

$$Outcome_{jt} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_j \cdot Post_t + Controls + \varepsilon_i$$

where  $Outcome_{jt}$  is the number of patents per capita (columns 1-3) and the number of patent attorneys per capita (column 4) around library  $j$  that is filed in year  $t$ .  $PatLib_j$  is an indicator if patent library  $j$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use a fixed effect for each combination of patent library, technology class and filing year in columns (1), (2) and (4). In column (2), we only use those observations where patent attorney data is available. In column (3), we control for the number of patent attorneys per capita. Data on patent attorneys stems from the historical rosters of registered patent attorneys at the USPTO. Standard errors are clustered on the patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure C-1: Aspirin

# UNITED STATES PATENT OFFICE.

FELIX HOFFMANN, OF ELBERFELD, GERMANY, ASSIGNOR TO THE FARBEN-FABRIKEN OF ELBERFELD COMPANY, OF NEW YORK.

## ACETYL SALICYLIC ACID.

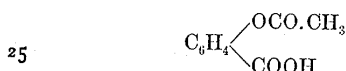
SPECIFICATION forming part of Letters Patent No. 644,077, dated February 27, 1900.

Application filed August 1, 1898. Serial No. 687,385. (Specimens.)

To all whom it may concern:

Be it known that I, FELIX HOFFMANN, doctor of philosophy, chemist, (assignor to the FARBENFABRIKEN OF ELBERFELD COMPANY, of New York,) residing at Elberfeld, Germany, have invented a new and useful Improvement in the Manufacture or Production of Acetyl Salicylic Acid; and I hereby declare the following to be a clear and exact description of my invention.

In the *Annalen der Chemie und Pharmacie*, Vol. 150, pages 11 and 12, Kraut has described that he obtained by the action of acetyl chlorid on salicylic acid a body which he thought to be acetyl salicylic acid. I have now found that on heating salicylic acid with acetic anhydride a body is obtained the properties of which are perfectly different from those of the body described by Kraut. According to my researches the body obtained by means of my new process is undoubtedly the real acetyl salicylic acid

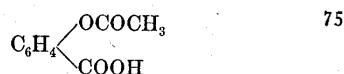


Therefore the compound described by Kraut cannot be the real acetyl salicylic acid, but is another compound. In the following I point out specifically the principal differences between my new compound and the body described by Kraut.

If the Kraut product is boiled even for a long while with water, (according to Kraut's statement,) acetic acid is not produced, while my new body when boiled with water is readily split up, acetic and salicylic acid being produced. The watery solution of the Kraut body shows the same behavior on the addition of a small quantity of ferric chlorid as a watery solution of salicylic acid when mixed with a small quantity of ferric chlorid—that is to say, it assumes a violet color. On the contrary, a watery solution of my new body when mixed with ferric chlorid does not assume a violet color. If a melted test portion of the Kraut body is allowed to cool, it begins to solidify (according to Kraut's statement) at from 118° to 118.5° centigrade, while a melted test portion of my product solidifies at about 70° centigrade. The melting-points of the two compounds cannot be compared, be-

cause Kraut does not give the melting-point of his compound. It follows from these details that the two compounds are absolutely different.

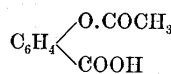
In producing my new compound I can proceed as follows, (without limiting myself to the particulars given:) A mixture prepared from fifty parts of salicylic acid and seventy-five parts of acetic anhydride is heated for about two hours at about 150° centigrade in a vessel provided with a reflux condenser. Thus a clear liquid is obtained, from which on cooling a crystalline mass is separated, which is the acetyl salicylic acid. It is freed from the acetic anhydride by pressing and then recrystallized from dry chloroform. The acid is thus obtained in the shape of glittering white needles melting at about 135° centigrade, which are easily soluble in benzene, alcohol, glacial acetic acid, and chloroform, but difficultly soluble in cold water. It has the formula



and exhibits therapeutical properties.

Having now described my invention and in what manner the same is to be performed, what I claim as new, and desire to secure by Letters Patent, is—

As a new article of manufacture the acetyl salicylic acid having the formula:



being when crystallized from dry chloroform in the shape of white glittering needles, easily soluble in benzene, alcohol and glacial acetic acid, difficultly soluble in cold water, being split by hot water into acetic acid and salicylic acid, melting at about 135° centigrade, substantially as hereinbefore described.

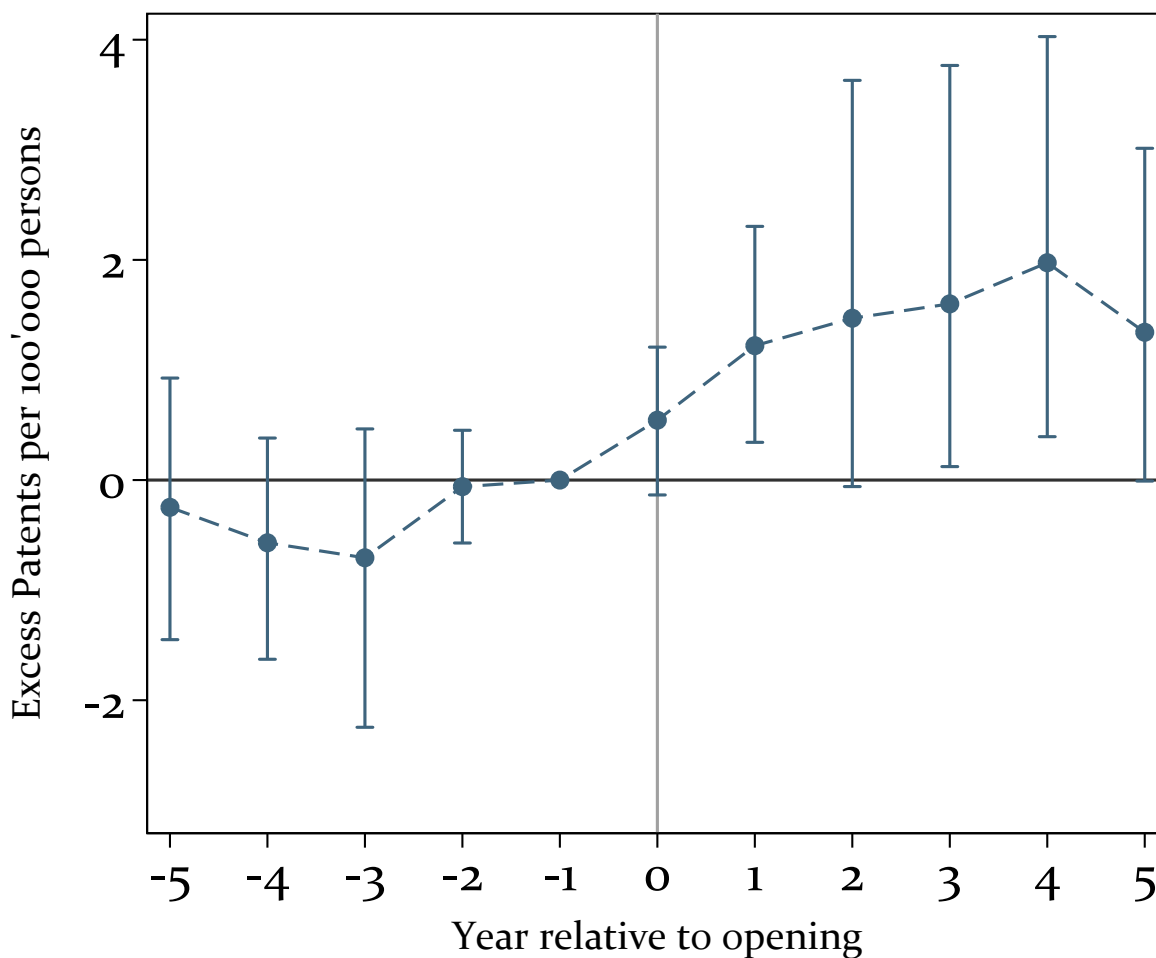
In testimony whereof I have signed my name in the presence of two subscribing witnesses.

FELIX HOFFMANN.

Witnesses:

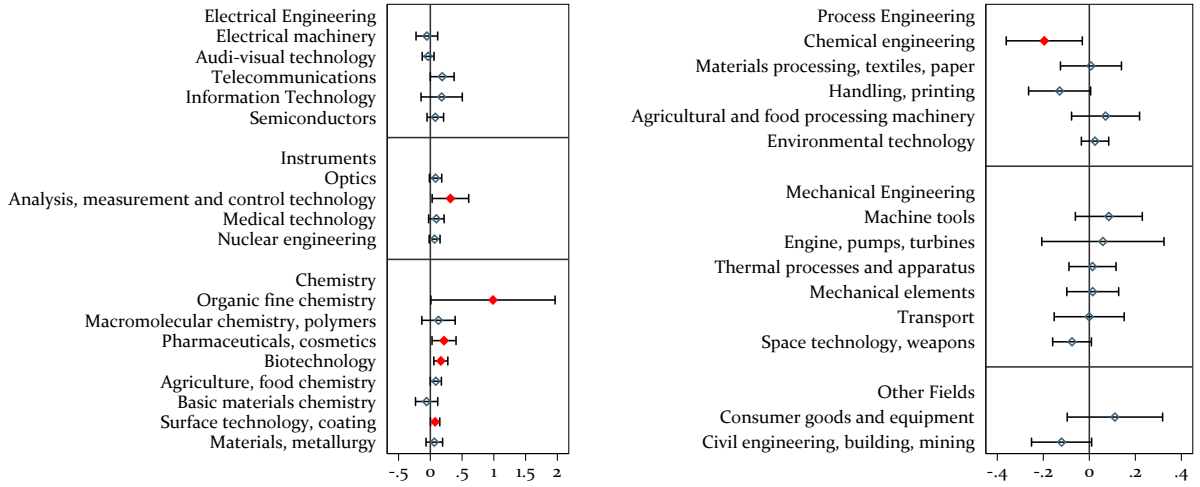
R. E. JAHN,  
OTTO KÖNIG.

Figure C-2: Non-parametric Evidence: Chemistry



**Note:** This figure shows the yearly average treatment effects on the treated of opening up a patent library on the average number of patents in chemistry within 15 miles of patent libraries relative to the average number of patents around matched federal depository libraries. The 95% confidence intervals are based on bootstrapped standard errors. We use the weights of Iacus *et al.* (2012) to arrive at the average treatment effect on the treated. We assign each patent library and all Federal Depository Libraries within the same state and within 250 miles as control group. We exclude the patent library of Burlington VT.

Figure C-3: Effect by Technology Category



**Note:** This figure 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:

$$\frac{\#Patents_{it\tau}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $\#Patents_{it\tau}$  is the number of patents within 15 miles of the library in year  $t$  and technological field  $\tau$ .  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. The technological fields follow the ISI-OST-INPI classification of 1995 as defined in Schmoch (2008). The range plots indicate the 90% confidence intervals for the coefficient that are plotted with a hollow diamond if the coefficient is not significantly different from zero or a full diamond if the coefficient is significantly different from zero. In Appendix C.3 we report the results for alternative classifications of technological sub fields.

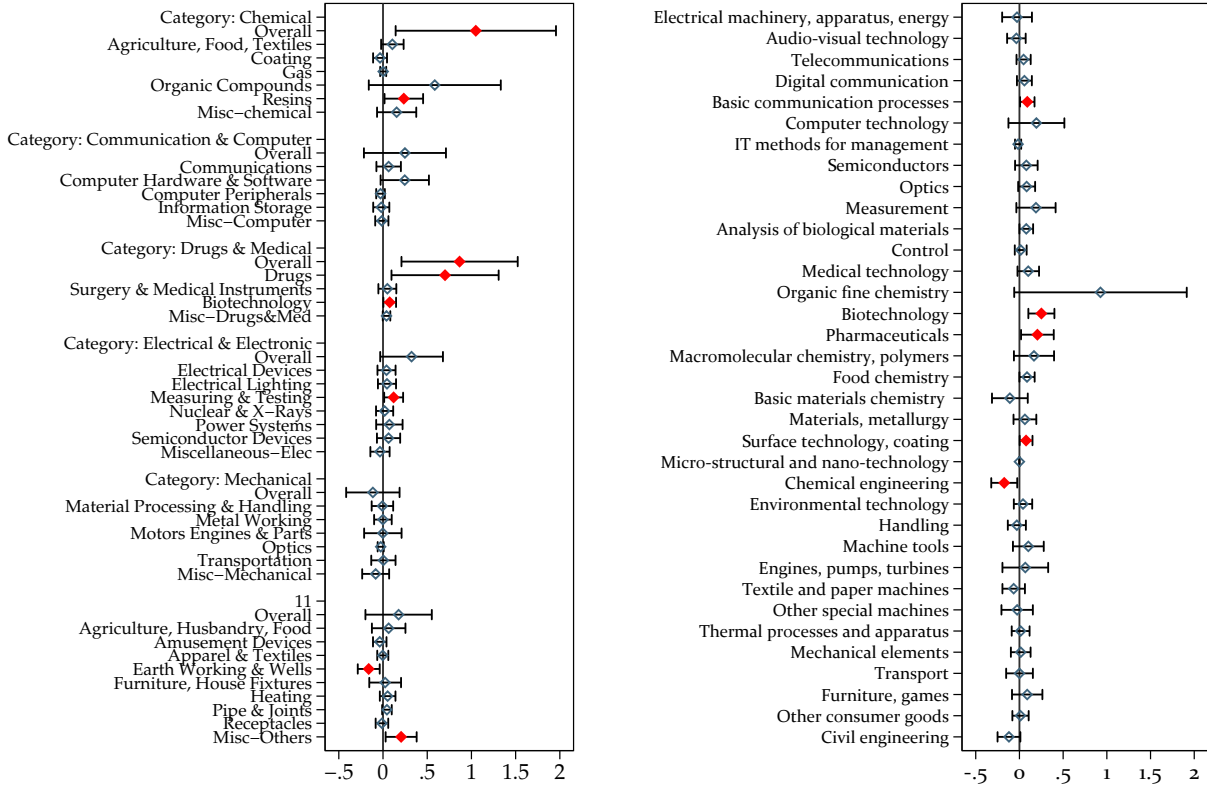
### C.3 Alternative Technology Classifications

In Figure C-4 we use two alternative technology classification to show the effects across fields. In Subfigure C-4a we use the NBER subcategory that are based on the USPTO technology classes. In Subfigure C-4b we use the 1995 version of the ISI-OST-INPI Technological Categories that are based on IPC classes. In both cases fields related to chemical and pharmaceutical drive the effect.

Figure C-4: Alternative Technology Classifications

(a) By NBER Subcategory (USPTO)

(b) ISI-OST-INPI Technological Categories 2008



**Note:** These figures 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:

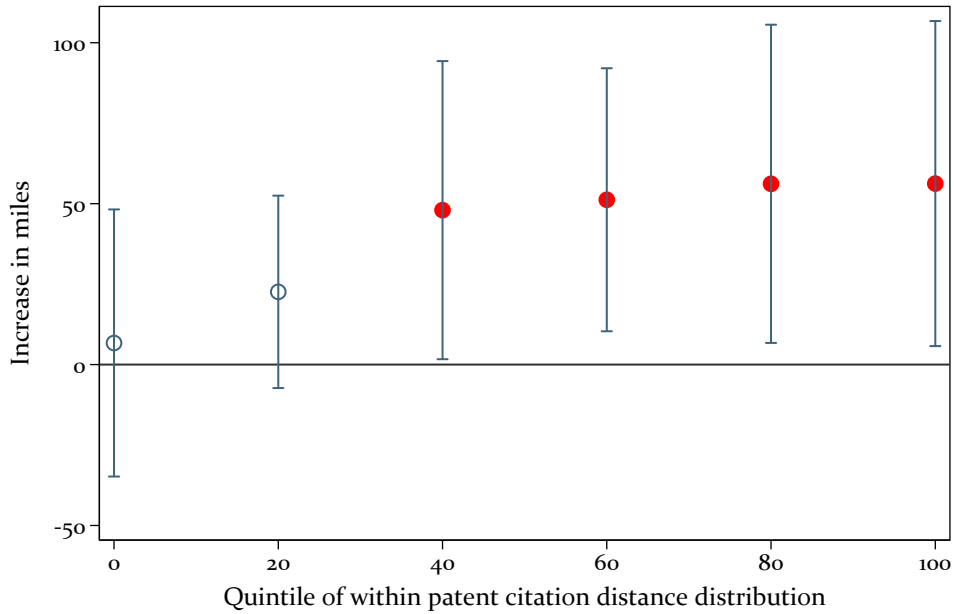
$$\frac{\#Patents_{it\tau}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $\#Patents_{it\tau}$  is the number of patents within 15 miles of the library in year  $t$  and technological field  $\tau$ .  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. The technological fields in Subfigure a) are defined following the NBER Subcategories of Hall *et al.* (2001) and in subfigure b) following the ISI-OST-INPI classification of Schmoch (2008). The range plots indicate the 90% confidence intervals for the coefficient that are plotted with a hollow diamond if the coefficient is not significantly different from zero or a full diamond if the coefficient is significantly different from zero. In Appendix C.3 we report the results for alternative classifications of technological subfields.

## C.4 Structure of Patents: Backward Distance Percentiles

In Figure C-5 we show the effect by percentiles of backward citation distance for young firms. There is a significant positive effect across almost the entire distribution.

Figure C-5: Effect on Backward Citations by Percentile, Young Firms



**Note:** This figure shows the increase in distance for each percentile of the within patent distance distribution for young firms. Young firms are those patenting for the first time no more than three years before the library opening.

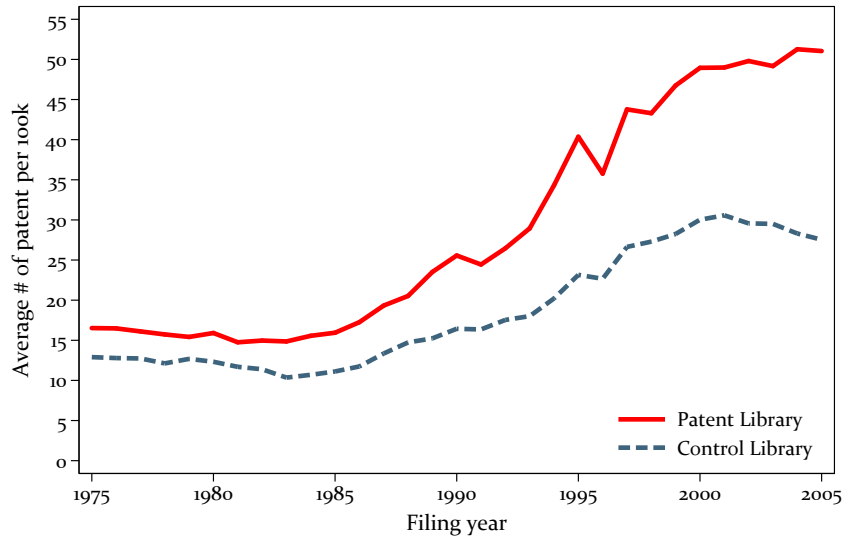
## C.5 Long Run Effects of Opening a Patent Library

While patent libraries opened in the Internet era did not have the same impact on patenting as those opened in earlier periods, it is possible that the impact of earlier patent libraries was, nonetheless, long-lived. For example, it is possible that library opening and the concomitant boost in regional innovation may have improved the overall environment for R&D and commercialization, attracting new innovators and, potentially supporting a longer-term increase in innovative capacity. Figure C-6 suggests that this, indeed, is the effect of patent library opening. It plots the average number of patents per 100,000 persons around patent and control libraries over time. To aid comparison we keep the sample constant over time, i.e., we include regions with patent libraries before they are opened. Patenting in the treated vs. control regions diverges significantly over time. The difference remains consistent and substantial beginning in the year 2000, although no new patent library is opened after 2001 and patents are freely available online during this time period. These results are consistent with the prospect that patent libraries provide a persistent boost to regions' innovation potential.<sup>34</sup>

<sup>34</sup>Note, that this difference in patent numbers is (at best) the upper bound of the effect of the patent library program. The effect in our main regression is identified under the assumption that nothing else changes at the same time that increases patenting and is correlated with the opening of the patent library.



Figure C-6: Averages Over Time



## C.6 Census Results: Robustness

To investigate whether the observed effects on industry dynamics depend on any particular patent library, we repeat our leave-one-out analysis for the main result on the entry of new establishments. Figure C-7 shows results when we drop libraries one by one. The coefficient does not change qualitatively, showing the robustness of our estimates.

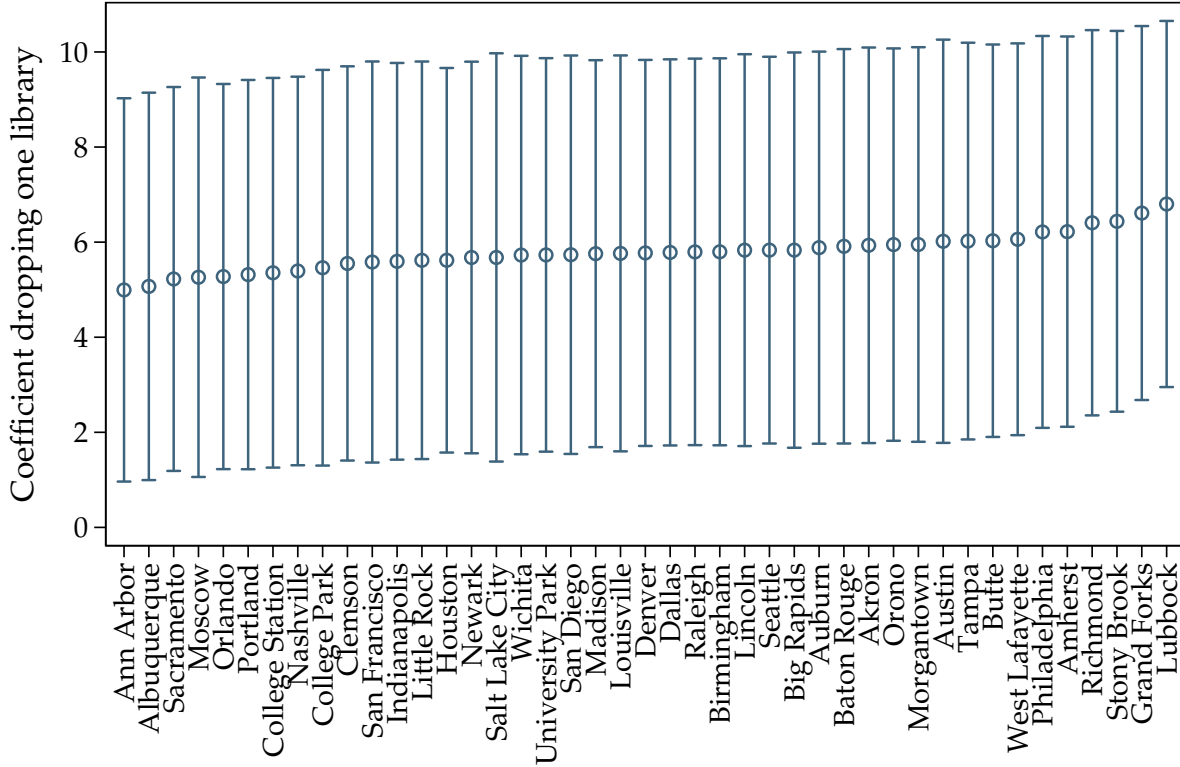
## C.7 Census Results: Selected Regression Results

Table C-1 reports the estimation results for selected outcomes of the difference-in-differences regressions underlying Figure 8.

---

This assumption is more credible in a short period before and after the opening of the patent library but less credible in the following 20 years. For example, large companies might reallocate their R&D to places that already have a cluster of inventors: Xerox PARC opened in Palo Alto in 1970 because there was already much research on computers in the Silicon Valley. Similarly, General Electric opened industrial labs in places with a strong knowledge base. Such relocations in space might reinforce the concentration of patents around patent libraries but they do not count toward the causal increase in innovation resulting from patent libraries.

Figure C-7: Stability: Leave-one-library-out Estimation



**Note:** This figure 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 sample are all establishments smaller than 10 employees. The estimation equation is:

$$\frac{Estab.Entry_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i$$

where  $Estab.Entry_{it}$  is the establishment entry variable from the census in the MSA of the patent library.  $PatLib_i$  is an indicator if the library  $i$  is a patent library and  $Post_t$  is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects.

Table C-1: Impact of Patent Libraries on Census Outcomes

	Small Companies				
	(1)	(2)	(3)	(4)	(5)
	Estab. Entry	Estab. Exit	Job Creation	Job Creation (Births)	Job Destr.
Post x Pat lib	5.8** (2.4)	2.1 (2.6)	28.8** (12.9)	22.4** (9.2)	7.4 (13.7)
Mean Dep.	176.79	149.59	1067.94	620.62	918.54
Obs.	3026	3026	3026	3026	3026
	Large Companies				
	(1)	(2)	(3)	(4)	(5)
	Estab. Entry	Estab. Exit	Job Creation	Job Creation (Births)	Job Destr.
Post x Pat lib	1.0 (1.0)	1.2 (0.8)	172.1 (107.1)	116.2 (79.6)	83.1 (67.0)
Mean Dep.	44.84	34.93	3828.18	1248.78	3289.34
Obs.	3887	3887	3887	3887	3887

**Note:** This table shows the results from difference-in-differences estimations with five years before opening as pre-period and five years after opening as post-period. In the upper panel we use small firms and in the lower panel large firms. Small companies are those with less than ten employees. Large companies are all companies with more than ten employees. As controls we use library and year fixed effects. Standard errors are clustered on the associated patent library level. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% levels, respectively.