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

Jeffrey L. Furman, Markus Nagler and Martin Watzinger*

March 6, 2018

How important is information disclosure through patents for innovation? To investigate this question, we examine the expansion of the USPTO Patent and Trademark Depository Library system from 1977 to 1997. While the exclusion rights associated with patents are national in scope, the opening of these patent libraries in a period before the internet yielded regional variation in the costs to access the technical information (prior art) disclosed in patent documents. Whereas the location of patent libraries prior to 1977 had been chosen by the USPTO, the timing and location of library opening in the 1977-1997 period was more random, as they were typically granted to the first qualified library in a state to request them. We find that after a patent library opens in a particular region, local patenting increases by 19% relative to control regions that have Federal Deposit Libraries but do not yet operate patent libraries. The facts that the response to patent libraries is greatest among young companies, that library opening induces local inventors to cite more geographically distant and more technologically diverse prior art, and that average patent citations received are not reduced by library opening, are consistent with the prospect that information disclosed in the patent documents is driving the paper's core findings. In additional analyses, we find that library opening is associated with an increase in local business formation and job creation. Taken together, our analyses provide evidence that the information disclosed in patent prior art plays an important role in supporting innovation. For managers, these results suggest that innovation supporting information spillovers are not only available through agglomeration,but may also be achieved by locating close to knowledge-enhancing institutions, such as patent libraries.

---

*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; Watzinger (corresponding author): martin.watzinger@econ.lmu.de. We thank Sharon Belenzon, Ben Jones, Ariel Dora Stern, Scott Stern, Bo Zhao, and participants at the NBER-AIEA Hong Kong, REER Atlanta, TILEC Tilburg, and various seminars for helpful comments and suggestions. We thank Jonas Kiral and Regina Seibel for excellent research assistance. We also thank 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. Furman gratefully acknowledges financial support from the NSF SciSIP grant, SES-1564368. Parts of this paper were written while Martin Watzinger was visiting Questrom School of Business at Boston University and Markus Nagler was visiting MIT Economics. 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

How does information disclosure through patents affect innovation? Disclosure of technical information is often invoked as one of the patent system's key economics justifications (e.g., Machlup and Penrose 1950; Scotchmer and Green 1990; Scotchmer 1991). In the legal debate, the U.S. Supreme Court has labelled disclosure the "quid pro quo of the right to exclude".[1] It is thought to both facilitate follow-on innovation by transmitting useful knowledge and by avoiding unnecessary duplicate innovation efforts.[2] However, intellectual property lawyers have recently started to doubt whether patent disclosure is indeed effective in fostering subsequent innovation for three reasons (e.g. Roin, 2005; Lemley, 2012). First, many patents are hard to read and it is difficult to extract information. It thus seems doubtful that useful information is transmitted. Second, many inventors do not even read patents because this increases the legal risk of "willful infringement" of prior art. And third, important inventions are often kept secret. Therefore, there are reasonable doubts about the usefulness of disclosed information through patents.

Understanding whether patent disclosure matters for follow-on innovation is of first order importance for the design of the patent system and our understanding of how to increase cumulative innovation. Yet, there is very little empirical evidence on how patent disclosure affects follow-on innovation because of a fundamental challenge for causal identification: the patent system makes the right to exclude competitors dependent on disclosing technical information which 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).

In this paper, we take advantage of the expansion of the USPTO Patent and Trademark Depository Library (PTDL or patent libraries) system from 1977 to 1997 to investigate the effect of

[1] Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470 (1974).

[2] For that reason, every patent law requires that a patent application must disclose a claimed invention in sufficient detail such that a person skilled in the art can carry out the claimed invention (Hall and Harhoff, 2012).

disclosure of patent information on innovation in the regions around newly-created patent libraries. While the exclusion rights associated with patents are national in scope, the opening of these patent libraries in a period before the internet yielded regional variation in the costs to access the technical information disclosed in patent documents.

This patent library system was created in the 1800s to provide patents and innovation-related resources for independent inventors, entrepreneurs, and incumbent firms. By 1977, 22 libraries had been established, primarily in New England and West of the Mississippi. Beginning in 1977, the USPTO embarked on an effort to open at least one patent library in each of the U.S. states. These libraries provided information on granted patents, including search tools, which were not otherwise available to inventors or attorneys outside of the USPTO headquarters location in Washington DC or in other patent libraries. Although we refer to the 'opening' of patent libraries, establishing Patent Deposit Libraries did nor require the construction of new facilities. Instead opening required that existing libraries dedicated sufficient space and staff for patent library materials.

Whereas the location of libraries prior to 1977 had been chosen by the USPTO, libraries after 1977 were granted with some degree of randomness within each state and in its time of opening, as they were typically given to the first qualified library in a state to request them. While some major cities in a state (e.g., Boston, MA) were more likely to have the capacity and demand for such institutions than smaller cities or towns (e.g., Springfield, MA or Worcester, MA), it is not clear whether innovation trends are more likely to have driven requests for such libraries in cities of similar sizes within states (e.g., Kansas City, MO vs. St. Louis, MO). While all nearly major innovation centers eventually receive such libraries, the specific timing with which cities and states received them appears random. For example, Dallas TX, and Denver CO received PDLs in 1977, as did Lincoln NE and Baton Rouge LA, while San Diego CA, Philadelphia PA, and San Francisco CA did not receive libraries until, 1984, 1986, and 1994, respectively. We leverage the randomness in location and timing of PDL opening to estimate the impact of disclosure on subsequent innovation.

In our main specification, we compare the number of patents in close vicinity around the patent library with the change in the number of patents around Federal Depository Libraries (FDLs). The 1252 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 serve first as FDLs and only later become patent libraries, making FDLs a natural control group. For each patent library, we use all Federal Depository Libraries that are located in the same state and within 250 km as a control group.

We find that, after a patent library opens, the number of patents within 25 km increases by 19%, an average of around three patents per year. This effect is highly localized and becomes insignificant at more than 25 km. Consistent with the prospect that increased access to patents is driving this effect, we find that patenting increases to a greater degree among young and small companies, which plausibly face larger barriers to access technical information than larger enterprises. The increase in patenting is most pronounced among patent libraries that are also university libraries. This result suggests that there is a complementarity between access to patent knowledge and technical education for the production of innovation.

We show that it is unlikely that concurrent shocks drive these effects. In the years before 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. Our results are robust to the use of alternative control groups, including regions with FDLs within 500km, regions only in the same state, and only ultimately-treated regions.

In additional analyses, we find that the effect is most pronounced in technology classes associated with chemistry and pharmaceuticals. 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). Looking at the effects over time, we find (a) that the effect is strongest after patent libraries introduced computer databases to search for prior art and (b) that it vanishes after the internet made patent literature universally available. This again speaks in favor of access to prior art as the driving force behind our findings and is at odds with diverging trends driving the results.

We also find that the structure of patents changes after a patent library opens: In particular, the distance to patents cited by inventors living close to a patent library increases and patents become more original. After a patent library opens, inventors thus start to work on problems that are less local and the geography of innovation becomes more dispersed. As well, we find evidence that patents start to cite different technological fields. There is no strong effect on patent quality as measured by forward citations. This is potentially important, as it suggests that the additional patents induced by PDL opening are not trivial patents but are similar in value to those granted prior to library opening. In summary, access to prior art facilitates the recombination of ideas across fields and helps the construction of an 'invisible college' of like-minded inventors building on each others' ideas.

To examine whether these effects exist in non-patent measures, we leverage Census data to assess the impact of patent libraries regional business dynamics. In our estimates, we find that Metropolitan Statistical Areas (MSAs) in which patent libraries were established experienced significantly greater rates of entry and exit of small firms than MSAs with comparable FDLs. In addition, we find that job creation increases in regions with patent libraries openings and that this is 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 affect local entrepreneurial environments.

This study is the first of which we are aware that demonstrates that access to technical information disclosed in patents can increase innovation. Disclosure is thought to be one of the key functions of the patent system. For example, Machlup (1958) argues that a patent, "serves to disseminate technological information, and that this accelerates the growth of productivity in the economy" (p. 76). Romer (1990) 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" (p. S84). However, critics have argued that the usefulness of disclosure through patents is limited (e.g., Roin, 2005; Lemley, 2012) and results from inventor surveys on 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).[3] Our study adds to this literature by demonstrating that increased access to prior art provided by patent libraries increases

[3]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; Hegde and Luo, 2017).

patenting for the subsample of small and young companies. Because access to patent documents does not affect the incentives to avoid willful infringement or the benefits of secrecy, we interpret our results as showing the impact of changes in the effective disclosure of patents on subsequent innovation.

More generally, our study contributes to the literature on research enhancing institutions by showing that investment in patent libraries helped to fuel regional innovation. Research enhancing institutions lower the costs of access to useful knowledge and thus help to foster geographical and intertemporal spillovers on which economic growth is based (Mokyr, 2002). 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 in particular in regions with libraries buying 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.

Finally, our results corroborate narratives about the role of patent depositories in the development of new inventions. For example, in the 1980s, Pfizer had attempted to develop a new microlide, an antibiotic of the same type as erythromycin, but with greater antimicrobial effect. Frustrated by its lack of progress, Pfizer was on the verge of closing down the program. While searching a patent database at the USPTO, Pfizer's chemists came across a patent belonging to the then-Yugoslavian firm Pliva the antibiotic azithromycin. Building on this patent, Pfizer developed Zithromax, which became the world's top selling antibiotic in the 1990s (Li, 2009). Jack Kilby, the co-inventor of the integrated circuit, is said to have read every patent document 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). Finally, 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 small and young inventors. For example, in his autobiography, the inventor Geoffrey Ball stresses the importance of technical information in prior patents and describes the Sunnyvale patent library as the "only place to research patents" (Ball, 2012).

Our results offer both general lessons for firm strategy in innovation-focused sectors and specific lessons for resource-constrained firms. While our analysis is focused on the benefits that accrue to firms in regions affected by patent libraries, the findings imply more broadly that access to technical information can offer a substantial and economically important benefit across firms. The fact that firms can benefit from published technical knowledge complements research documenting that investments in internal knowledge capabilities redound to firm benefit (Cohen and Levinthal, 1989). As well, our findings demonstrate that resource-constrained firms, such young and small ventures, can benefit from disproportionately from public investments in research-enhancing institutions and imply that these firms may leverage such knowledge hubs to achieve information and innovation benefits without relocating in high-density knowledge clusters.

The remainder of this paper is organized as follows. Section 2 describes the US Patent Depository Library Program and the US 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 auxiliary results to examine the underlying mechanism. Section 6 concludes.

2 The U.S. Patent Depository Library Program[4]

The nationwide network of Patent and Trademark Depository Libraries (PTDLs or patent libraries) traces its beginning to the year 1871 when the United States Patent and Trademark Office (USPTO), then known as The Patent Office, first started distributing copies of patents to a small number of libraries. Until then, official patent documents were housed and available for widespread perusal at only one location, at The Patent Office, in Washington, DC.[5]

In 1871 the first eight libraries received patent documents: The New York State Library, the Boston Public Library, The Public Library of Cincinnati and Hamilton County, the Science and Engineering Library at Ohio State University, the Detroit Public Library, the Los Angeles Public Library, the New York Public Library, and The St. Louis Public Library. By 1976, the number of patent libraries had grown to 22, most of which were located in the industrial Midwest and eastern seaboard. Each library received weekly deliverables of unbound paper patents, the Official Gazette

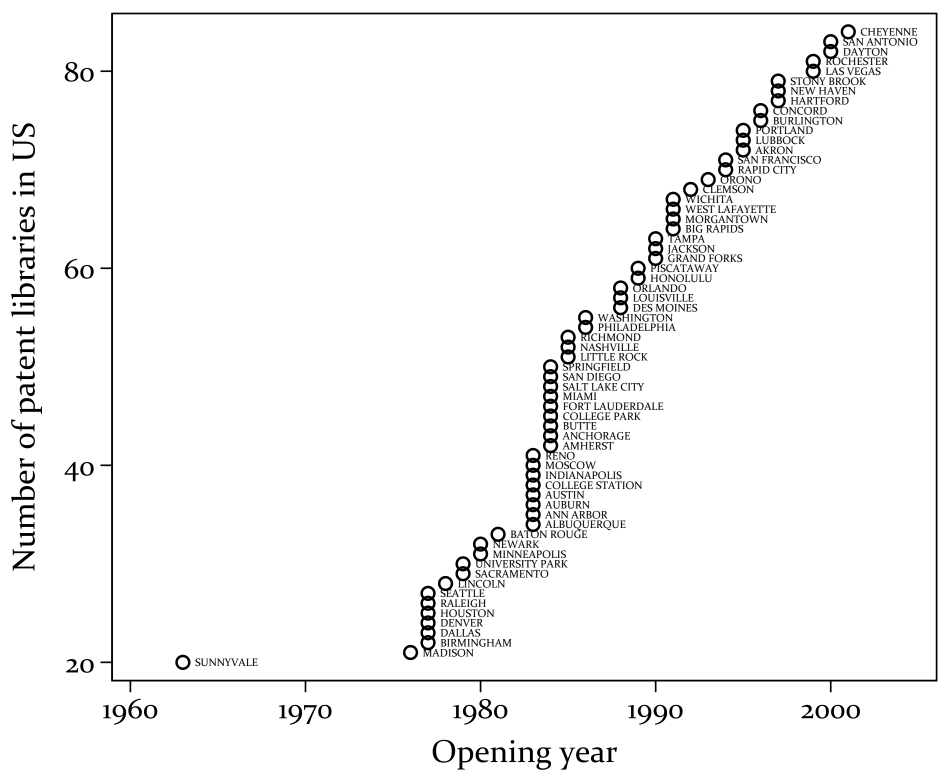
[4]The history section follows the descriptions in Sneed (1998) and Jenda (2005).

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

of the U.S. Patent and Trademark Office, and two search indices.

After 100 years of relative inactivity, the USPTO began an aggressive expansion of the patent library system in 1976-1977 with the aim of increasing the number of patent libraries by at least three per year and, ultimately, operating at least one patent library in each state.[6] This latter aim was achieved two decades later in 1997. The map in Figure 1a identifies the 22 libraries in operation before 1976, while Figure 1b lists all patent libraries opened after 1976. Figure 2 shows the expansion of the patent depository library system over time.[7] Currently, about half of the membership consists of academic libraries with nearly as many public libraries.[8] 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).

Figure 2: The expansion of the patent depository library program



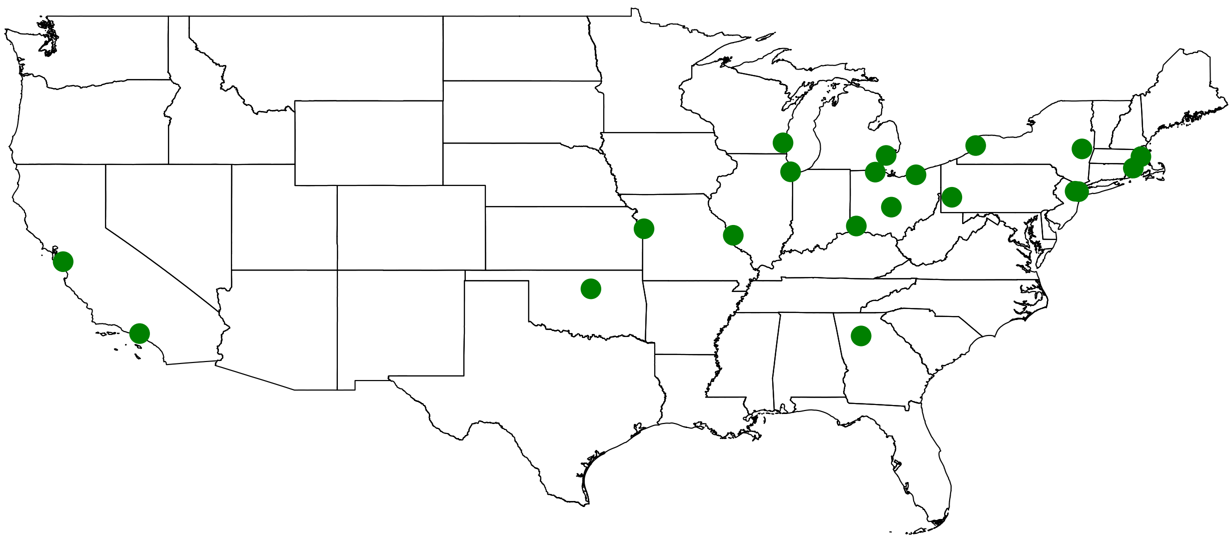
[6]The reason for this change in policy was the leadership of USPTO assistant commissioner William I. Merkin who started to evaluate the patent library system in 1974. This evaluation led its overhaul starting in 1977.

[7]Table 6 and Table 7 in the appendix list patent libraries up to 2002.

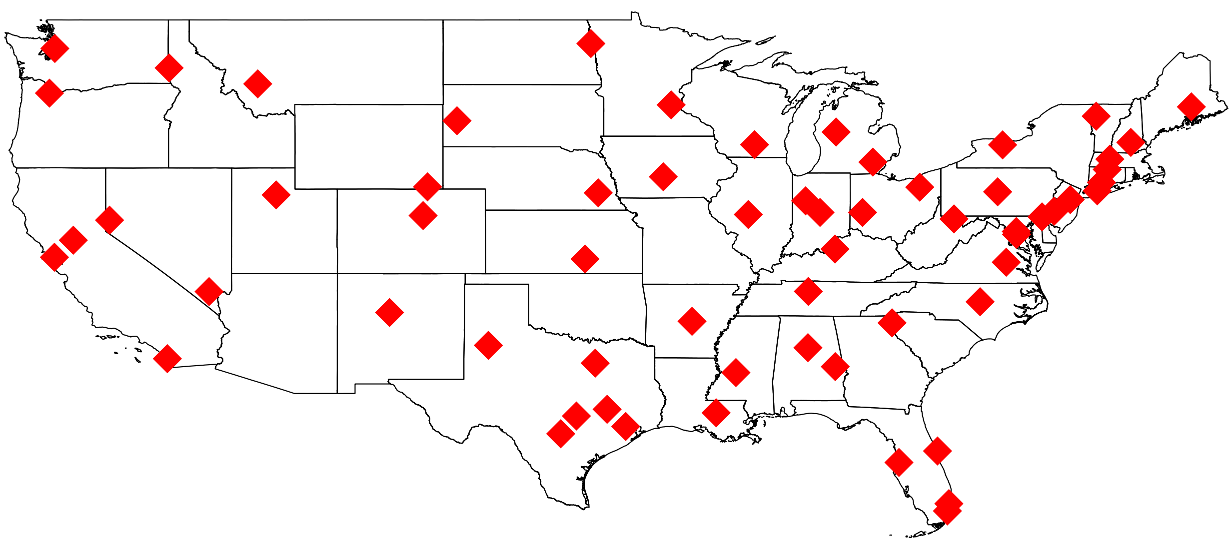
[8]Since 1871, six PTDLs have withdrawn for various reasons, including library closing, no funding for the back file, and a change in institutional priority creating a lack of ability to perform required services.

Figure 1: Location of all patent libraries in the US

(a) From 1870 to 1976



(b) From 1977 to today



Note: Figure 1a shows the position of patent libraries in the continental United States opened before the major expansion in 1977. In Figure 1b diamonds indicate the location of patent libraries opened in or after 1977.

Beginning in 1976, the patent library system was not only expanded but also reorganized: Libraries were eligible to apply to become patent libraries if they fulfilled a number of requirements. First, they 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. 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 every annual PTDL Training Seminar in Washington, DC.[9] Third, they had to provide free public access and a collection of search tools for the public.

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

According to the librarians we interviewed, most applicant patent libraries ad had prior previous experience handling government documents as Federal Depository Libraries before applying to become Patent Deposit Libraries. Federal Depository Libraries make U.S. federal government publications available to the public at no cost. As of 2008, there were 1252 Federal Depository Libraries, at least two in each of the 435 Congressional Districts.[10]The librarians' qualitative reports are consistent with our data. Eighty four percent of the patent libraries in our sample are also part of the Federal Depository Library Program. Because of this structure and the requirements associated with serving as in either library program, Federal Depository Libraries constitute a natural control group for Patent Depository Libraries.

Interestingly, the USPTO continued 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 Cen-

[9]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.

[10]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.

ters, continue to offer value to the communities they serve by offering assistance with patentability searches and databases that offer more sophisticated prior art search capabilities than publicly-available resources.

3 Data and Empirical Setup

For our empirical analysis, we combine data on libraries with geolocated patent data and population data from the U.S. census.[11] 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.[12] Patent data is from PATSTAT and we compute backward and forward citations along with measures for the generality and originality for each patent (Hall *et al.*, 2001). For the geolocation of the inventors and inventor disambiguation we rely on the data of Li *et al.* (2014) for the publication years of 1975 to 2005 and of Balsmeier *et al.* (2017) for the years from 2005 to 2010. For the geolocation of patents with a publication year before 1975 we use the data of Petralia *et al.* (2016).[13] If there are several inventors on a patent we give each location a share of the patent.[14]

Figure 3a plots the patent libraries together with patent data across space. To aid visualization we also plot the centroid of each county in light grey and 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 within 25 km of the new library by the population in the area. We use the U.S. Census data for incorporated places at the end of the sample period in 2010.

To construct a counterfactual for regions with patent libraries, we select a control group consisting of regions with Federal Depository Libraries (FDLs) that are within the same state as the treated patent library while also being a minimum of 25k and a maximum of 250km away from the

[11] In Appendix 7 we describe the data assembly process in a step-by-step manner.

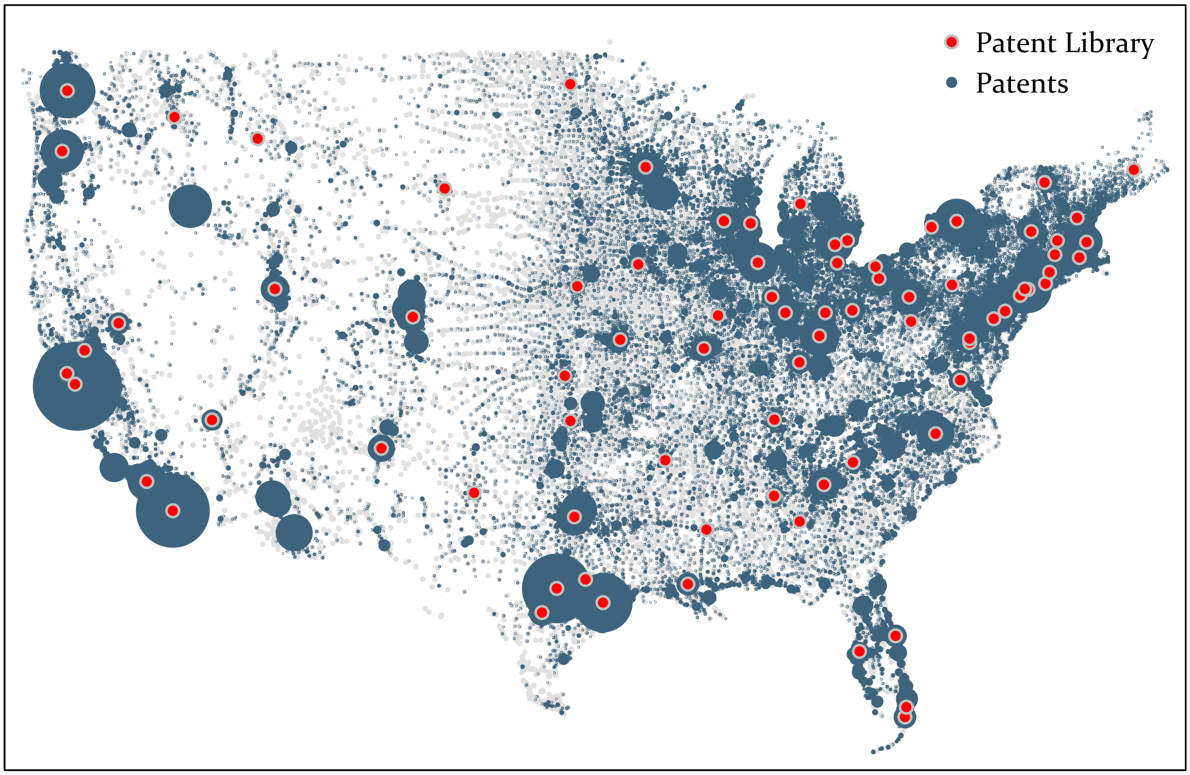
[12] The Federal Depository Library Directory is available on https://catalog.gpo.gov/fdlpdir/FDLPdir.jsp (last accessed 2017-07-30).

[13] In unreported regressions we use the alternative dataset of geolocated patents of Morrison *et al.* (2017) and find the same results.

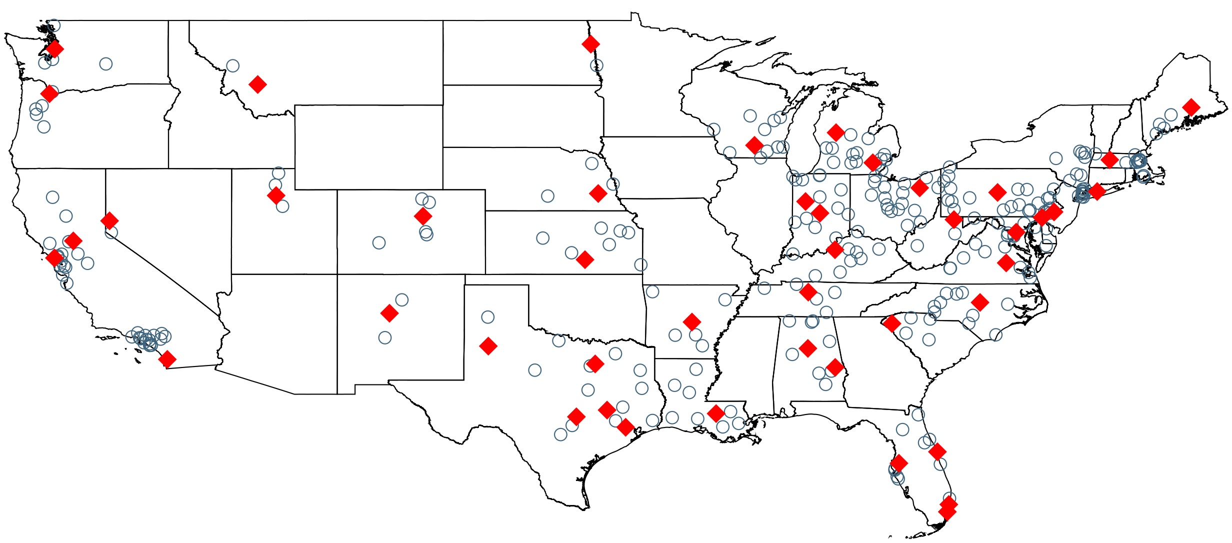
[14] To use patents as a measure for innovative output is standard but not uncontroversial. In our particular case, patent libraries also might increase patenting without increasing innovation because they might make it easier to file a patent or because the librarians might give advice on how to structure a patent. Yet, this seems unlikely because a US patent application can be mailed from any post office and the employees of patent libraries are only allowed to help with the search for prior art but not with the preparation of a patent filing.

Figure 3: Patents and Libraries across space

(a) # Patents and Patent Libraries



(b) Patent and Control Libraries



Note: In Figure 3a, 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 US county in light grey. In Figure 3b, the red dots show the position of patent libraries. The hollow dots show the positions of control libraries.

Table 1: Sample selection

Sample	Patent Libs	FDLs	Libraries dropped
All patent libraries	84	665	
Only Patent Libraries after 1975	64	521	
Only Patent Libraries that are also FDL	53	380	Cheyenne (WY), Concord (NH), Des Moines (IA), Hartford (CT), Jackson (MS), Minneapolis (MN), New Haven (CT) Piscataway(NJ), San Antonio (TX), Springfield (IL)
No control library available	50	380	Las Vegas (NV), Moscow (ID), Washington (DC)
Estimation sample	48	368	Burlington (VM) and Rochester (NY)

treated library. At the moment, there are 1251 Federal Depository Libraries in the United States, including at least two in every Congressional District. Most patent libraries were first Federal Deposit Libraries before designation as patent libraries. These are the libraries more likely to possess the space, human capital, and library infrastructure to become Patent Deposit Libraries. Moreover, as our data later show, FDLs operate in regions whose characteristics are similar to those of patent libraries.

To arrive at our final sample we drop all patent libraries that were opened before 1975, all patent libraries that are not Federal Depository Libraries and all libraries without a control between 25km and 250 km. We also drop the libraries of Rochester in the state of New York and Burlington in Vermont. The reason is that both have an extremely high patent per capita ratio because they host Kodak, Xerox, and Bausch & Lomb in case of Rochester and IBM in case of Burlington. As a consequence, there is no suitable control region for these two libraries. We thus arrive at 48 patent libraries that opened after 1975 along with 368 control libraries. In Table 1 we show how each of these selection steps influences our sample. Figure 3b shows the position of all patent libraries and all Federal Depository Libraries in our sample.

Table 2 shows summary statistics for patent libraries and matched Federal Depository Libraries in the year before the opening of the patent library. The number of patents per capita, of young, and of small firms are similar prior to the opening of the patent library. The only statistically significant

Table 2: Summary statistics in the year before opening

	Patent Libraries	Control Libraries	Diff	P-Value
# Patents/100k	17.60	13.46	-4.14	0.10
# Patents	130.67	118.44	-12.23	0.57
# Pat. small firms/100k	8.22	5.86	-2.35	0.06
# Pat. big firms/100k	9.38	7.60	-1.78	0.31
# Pat. young firms/100k	0.16	0.16	-0.00	1.00
# Patents old firms/100k	17.44	13.30	-4.14	0.10
Population in 100k	7.77	12.78	5.01	0.00
Uni Library	0.72	0.73	0.01	0.91

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

difference is that within 25 km the population size is significantly larger for control than for patent libraries. The reason for this is that the inclusion of New York in the sample of cities containing a Federal Depository Library inflates the mean for the control sample. If we consider median rather than mean population, regions with patent libraries are larger than those with Federal Depository Libraries.[15]

In our empirical analysis, we estimate the effect of opening a patent library on innovation within 25 km (15.5 miles) around the new library. We can interpret our estimates as causal if, in the absence of the opening of the patent library, the number of patents per capita in this region would have had the same trend as the number of patents around the control libraries. One potential concern about this identification assumption is that a libraries may have applied to become Patent Depository Libraries because their librarians or administrators or the USPTO had expected that innovative activities in these regions will ripe for immediate development.

While we cannot rule out this possibility entirely, we believe that it is not a substantial threat to our analysis. One factor in our consideration is that we are considering a period of program expansion in which the USPTO had the aim of opening up a library in every state. This program

[15] 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 show patenting activity in some years. In Section 8.1 in the Appendix, we show that balancing improves and all results hold when we drop these observations. 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.

was motivated by equitable access across states and, thus, seems less likely to be endogenous to local economic conditions. We also interviewed several librarians and the reasons to become a patent library seemed idiosyncratic. One librarian argued that status consideration between libraries played a role, another said that the librarian in charge wanted to take part in the seminars in Washington DC, and still others cited public service considerations. While most centers of innovation ultimately receive patent libraries, the key to our identification strategy is the timing with which the libraries are opened must be random with respect to innovation trends. Although nearly all major innovation centers eventually receive such libraries, the particular dates on which libraries open does not follow a pattern of increasing or decreasing innovation importance, either in levels or in changes.

To show that the assumptions underlying our identification strategy are reasonable, in Subsection 4.2 we conduct several robustness checks. First, we show that before the patent library was opened, 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 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 show that our results are robust.

4 Results

The opening of a patent library potentially reduces the costs of searching for prior art for inventors in close proximity. In this section we estimate the effect of the opening of a patent library on the number of patents within 25 km around the patent library.

It is a priori unclear if there is an effect of opening up a patent library on local innovation. On the one hand the improved access to the patent literature might increase innovation if inventors read patent literature and draw valuable information from it (Machlup and Penrose, 1950; Scotchmer and Green, 1990; Scotchmer, 1991; Landes and Posner, 2003). On the other hand, a number of legal scholars claim that patent disclosure does not work because patents are opaquely written and helpful inventions are kept secret from competitors (Levin *et al.*, 1987; Moser, 2011, 2013). Even more, inventors who read patents open themselves for willful infringement lawsuits and therefore might ignore the patent literature (Roin, 2005; Lee and Cogswell III, 2004).

4.1 Patent Libraries Increase Local Innovation

We start by investigating whether opening a patent library has any impact on patenting within 25 km around the new library. This distance corresponds to the approximate commuting distance of two-thirds of American commuters at the conclusion of the study period USD (2003)and serves as a convenient radius around which to measure the potential impact of patent libraries. We explore the robustness of the analysis to alternative distance measures. In Figure 4, we plot the yearly difference in the number of patents in the 25 km radius around our control and treated samples. For each library region, the data in the figure incorporate patents applied for in the five years before and the five years after the library opening, normalizing the measure around zero in the year of initial operation. 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).

We find that the number of patents around the patent library increases significantly in the year after opening and is stable in the following five years. The yearly difference is significantly different from zero at the 10% level. 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 confidence in our interpreting the estimated effects as causal. In Appendix 8.2 we compare simple averages and find the same result.

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

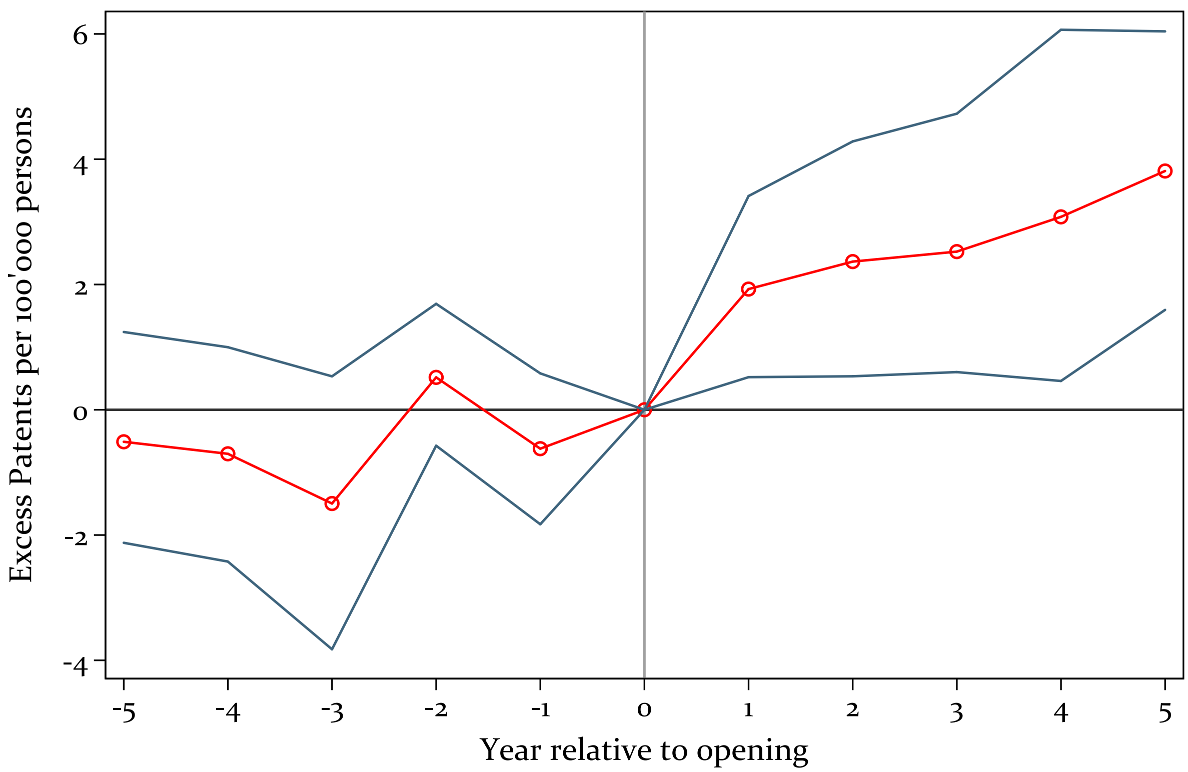
$$\frac{\#Patents_{it}}{Population_i} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\ FE + Year\ FE + \varepsilon_i \qquad (1)$$

where i indexes each library (i.e., the 25 km region around each library), $PatentLib_i$ 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 the years following patent library opening. We incorporate both library and year fixed effects as controls.[16] The coefficient of interest, β_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

[16]The baseline effect of $PatentLib_i$ is taken up by the library fixed effects.

Figure 4: 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 25 km of patent libraries relative to the average number of patents around matched federal depository libraries. The 90% confidence intervals are based on bootstrapped standard errors. We 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 Library within the same state and within 250 km as control group. We exclude the patent libraries of Burlington and Rochester.

Table 3: Patent libraries and local innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Baseline		Size		Age		University	
	250km	Citations	Dollars	Small	Large	Young	Old	Yes	No
Post	-0.6	-20.0	-4.4	-0.5	-0.1	2.8***	-3.4***	-0.9	-0.7
	(0.6)	(21.2)	(3.4)	(0.4)	(0.4)	(0.3)	(0.6)	(0.7)	(1.0)
Pat Lib x Post	2.9**	56.3***	11.1***	1.3*	1.6	1.9**	0.9	3.6**	1.0
	(1.3)	(18.8)	(3.5)	(0.7)	(1.1)	(0.9)	(1.1)	(1.7)	(1.3)
Mean Dep.	15.3	209.9	59.1	7.2	8.1	3.1	12.2	14.9	16.4
R2 (within)	0.10	0.11	0.15	0.15	0.04	0.43	0.09	0.10	0.20
Obs.	4576	4576	4576	4576	4576	4576	4576	3322	1254

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 $PatentLib_i$ 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 250km as controls. In column (2) we weigh each patent with its forward citations. In column (3) we use the patent values from Kogan *et al.* (2017) to weigh each patent with its dollar value. We windsorize the Dollar values at the 90th percentile to adjust for outliers. In column (4) and (5) 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 the following two columns we split the dependent variable by young and old assignees. An assignee is young if it filed its first patent after the opening of the library and old otherwise. 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.

of patents per capita in close vicinity of the patent library increased on average by 2.9 relative to the control group. This is an increase of 19% relative to the average. The effect is of similar magnitude in percentage terms if we weight the patents by the number of forward citations, as in column (3) or by the average value of a patent in the same technology class and the same publication year, as in column (3). The number of citation-weighted patents increases by 27% while the number of dollar-weighted patents increases by 19%.

We use a back-of-the-envelope calculation to approximate the economic impact of the additional patents on the local area. Specifically, we weight each patent by the value estimated by Kogan *et al.* (2017), using the results in column (3). The Kogan *et al.* (2017) weightings imply that the average region experiences a boost in patenting whose value is approximately $11 million per year. 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 the fact 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.

Our further results suggest that young and small companies play an important role in the boost in patenting. In our analysis, we consider a company to be small if it has fewer than 20 patents at the data of local patent library opening. We consider firms to be young if their first patent had not applied for a patent in the years prior to the library opening. 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 patents library increases the number of patents by small companies and young companies by 18% and 60%, respectively, relative to the (low) mean. The effects for large and old companies, reported in columns (5) and (7), are not statistically different from zero, although the point estimate for large firms is similar in magnitude to that for small firms.

In columns (8) and (9), 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 8.3 we demonstrate that the effect is driven by firms rather than other types of assignees and there is little difference between regions with historically high or low patenting levels.[17]

4.2 Auxiliary Analysis and Robustness

The main concern for our estimation strategy is that regions with a new patent library may have experienced equal patent increases even in the absence of patent deposit library dedication. This could occur, for example, if the USPTO were to choose to place patent libraries in regions in times when they were expected to achieve innovation spurts. An analysis of patenting under such circumstances coincidence would yield findings similar to those reported above, they would not coincident with but not induced by library opening. In this section we report the results from auxiliary analyses that shed light on this possibility, although our conclusion is that such a coincidence is not likely. We report the results of these additional analyses in Table 4, comparing new specifications to our preferred specification, which is reported in column (1).

Pseudo Outcomes: Effect is Not Measurable Before Patent Libraries Opened and Outside of its Close Vicinity

Figure 4 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 be receive to a patent library and those regions with Federal Depository Libraries that do not obtain patent libraries. The lack of differences in patenting flows prior to the patent library opening suggests that the parallel trend assumption holds. The number of patent applications that are ultimately granted increases in the years immediately after patent library opening. Were patent libraries to be opened based on the expectation of regional innovation booms, the timing of the opening and the expected increase would have to coincide perfectly. Particularly considering the randomness in library locations opened in particular years, USPTO prescience regarding next-year innovation booms seems like a less likely prospect than the possibility that access to patent technical information supports the translation of ideas into patent applications beginning in the year following library opening.

One other feature of the results that suggests that local access to technical information is the

[17] In Section 8.6 in the Appendix, we also show that relative to a "synthetic development" of PTDL regions holding their share among all U.S. patents constant, libraries increase local innovation.

Table 4: Auxiliary results

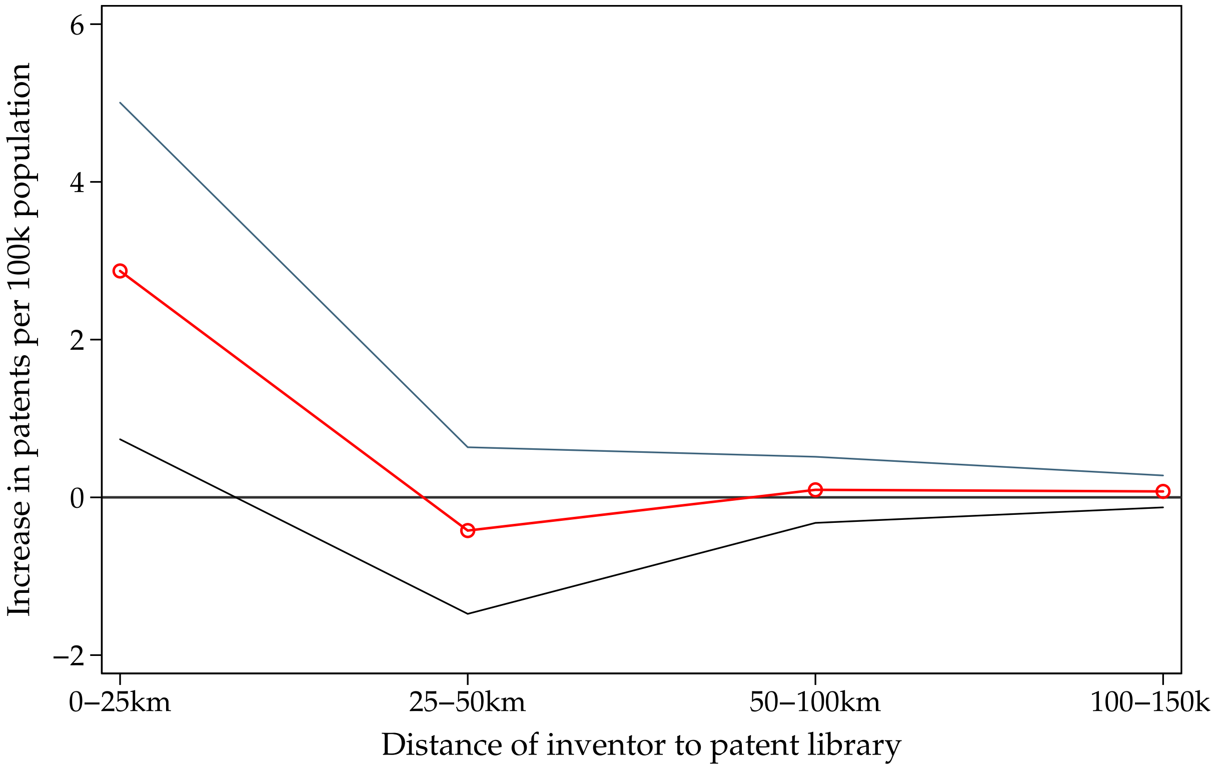
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Base-line	Patents between	Fake Opening	Matched by		Control group			All libraries
	250km	25-50km	250km	Pop	Patents	500km	State	Within	250km
Post	-0.6	0.2	-0.3	-2.3	-1.7	-0.4	-0.4		-0.9
	(0.6)	(0.7)	(0.5)	(1.5)	(1.4)	(0.5)	(0.5)		(0.6)
Pat Lib x Post	2.9**	-0.4	0.5	5.9*	4.7*	2.6**	2.6**	6.6*	5.8
	(1.3)	(0.6)	(1.5)	(3.3)	(2.7)	(1.2)	(1.2)	(3.7)	(4.1)
Mean Dep.	15.3	6.0	16.6	17.5	16.0	14.5	14.5	23.3	18.6
R2 (within)	0.10	0.06	0.13	0.11	0.13	0.11	0.11	0.33	0.10
Obs.	4576	4576	4722	1089	1595	5247	5247	2970	4730

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:

$$\frac{\#Patents_{it}}{Population} = \alpha + \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Controls + \varepsilon_i$$

where $PatentLib_i$ 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 250km as controls. In column (2) we use the number of patents between 25 and 50km as outcomes. In column (3) we assign a treatment indicator to the FDL closest to the patent library and drop patent libraries from the sample. In column (4) we use FDLs within 250 km and match additionally on similar sized population within 25 km of the treatment and control library. We define a city with a control library as similar if it is less than three times the size of city with the patent library. In column (5) we include only FDLs which had in their close vicinity not more than 50% more filed patents in the year before the opening than patent libraries. In column (6) we extend the control sample to FDLs within 500 km and in column (7) to all FDLs within each state. In column (8) we use not-yet opened patent libraries as control for patent libraries. In column (9) we additionally include Burlington, Vermont and Rochester in New York in our estimation sample. We use the weights suggested by Iacus *et al.* (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the (assigned) patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure 5: Effect of patent libraries by distance



Note: This figure shows the coefficient β_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. $PatentLib_i$ 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.

root of the findings is the absence of the effect for more distant inventors. Column (2) of Table 4 and Figure 5 show that the increase in patents is localized in a small area around the patent library. For patents localized more than 25 km away we do not find any effect. To arrive at these results, we use the number of patents in different distance bands around the treatment and the control library as outcome. Our result implies that the number of patents only increases around the patent library but not in the wider area. If a region is chosen to get a patent library based on an expected increase in innovation the government must pick exactly the right spot where patenting will increase.

SUTVA: No Effect of Pseudo Treatment of Closest Control Library

Another prospect that we should explore regards the possibility that opening up a patent library causes inventors to move in space, though not necessarily increase innovation overall. If inventors move to the location of patent libraries, our results might simple reflect a change in the spatial distribution of patents and not 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 patents around the control libraries. If inventors move towards newly-established patent libraries, it is reasonable to assume that they would move most from the closest comparable area. To test this possibility, we re-run our analysis omitting patent libraries and, instead, assigning instead 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 an substantial magnitude suggests that there is no differential trend between closer control libraries and libraries that are further away. Therefore, interference does not appear to be a concern in this study.

Robustness to Alternative Control Groups

Our estimation strategy thus far has been based on the assumption that Federal Depository Libraries in the same state and within 250 km are a suitable control group for patent libraries. In columns (4) to (8) we relax this assumption and explore the robustness of the results to the use of alternative control groups. In columns (4) and (5) we choose more narrow control groups. In the analysis in column (4) we match using only those Federal Depository Libraries in regions whose population between 50% and 200% of the city of the population in the region of the treated patent library. In column (5) we include only those Federal Depository Libraries whose regions achieve patenting levels between 50% and 200% of those of the treated patent libraries in the year before patent library opening. In each case, the choice to narrow the control sample yields a larger estimated effect, though at the cost of less precision in the estimate. In columns (6) and (7) we compare treated library regions to more broadly-defined control groups. In column (6) we include all Federal Depository Libraries within 500km (as opposed to 250km, as in our preferred specification) and in column (7), we include all federal depository libraries within the state (regardless of distance from the patent library). The coefficients are in both cases of similar size and statistically significant.

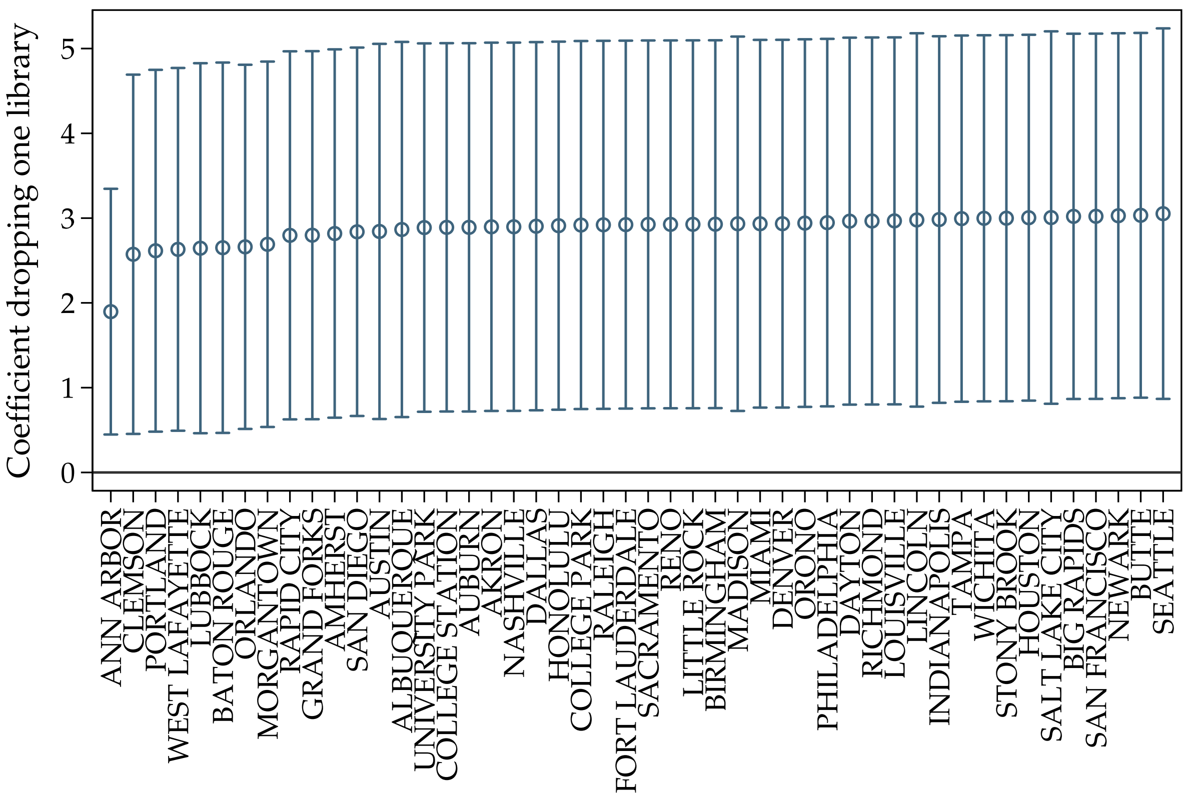
In column (8), we report the results of a model that identifies the impact of patent library opening using only those libraries that ultimately receive patent collections. In some sense, this is our strictest test, the identification here relies only on the timing of the treatment and not on any differences between libraries that receive patent collections and those that do not. For this analysis, we include not just the five years prior to patent library opening as the pre-period, but all years for which data are available. This choice does not materially affect the coefficient in the regression, but does increase the power of the analysis. The results of this model are consistent with the findings that use the Federal Depository Libraries as controls. Relative to the period prior to patent collections becoming available, regions that receive patent deposit collections experience a statistically significant and economically meaningful boost in patents. The coefficient on Patent Library * Post in is larger for the within sample, but less precisely estimated than in prior models.

In our final set of robustness checks, we explore the role of specific libraries in our results. In Figure 6 we run our main analysis, dropping individual libraries 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 2.9 to around 2.0, while making the estimate more precise but still within the initial confidence interval. As we described above, our main sample excludes the patent libraries of Rochester NY and Burlington VT. Each of these libraries' regions has an extremely high patent per capita ratio because Rochester was the headquarters for Kodak, Xerox, and Bausch & Lomb and Burlington was the home of IBM's major research facility. Each constitutes a substantial innovation outlier in its local area. As a result, we could not identify a control region within 250 km and within the same state that achieved even remotely similar levels of per capita patenting. In column (9) we add these two libraries into our main analysis. We find an post library opening effect size greater than that in our preferred specification, but also that the additional noise renders the coefficient indistinguishable from zero.

5 Exploring Potential Mechanisms

Our prior analyses document that the opening of patent libraries induces an increase in local patenting. In this section we explore mechanisms that help explain these findings. The first three analyses

Figure 6: Stability: Leave-one library out estimator



Note: This figure shows the coefficient β_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:

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

where $\#Patents_{it}$ is the number of patents within 25km of the library in year t. $PatentLib_i$ 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.

are consistent with the possibility that the technical information available in patent prior art plays a key role in enabling local businesses to benefit from patent materials. The latter analysis demonstrates that the impact of libraries is not limited to patenting, but also induces entrepreneurship and associated business dynamics.

5.1 Technology: Effect is Concentrated in Chemistry

Our first analysis examines which industries account for the patenting response. Figure 7 reports the results of estimating Equation 1 using patents by technological field as our outcome variable. To define technology fields we use a classification that aggregates IPC technology classes to larger sub fields (Schmoch, 2008).

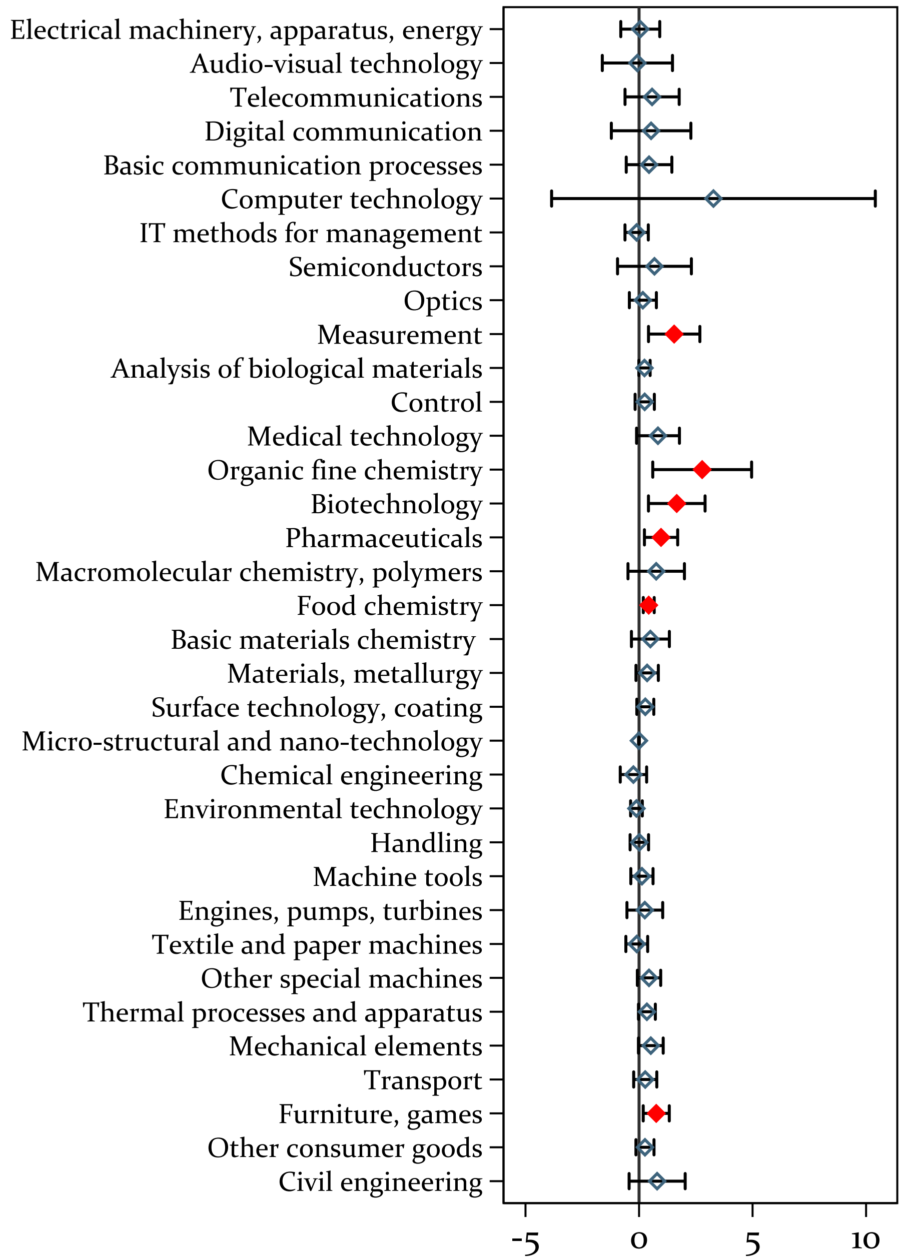
We find that the increase in patents is most pronounced in fields related to chemistry and pharmaceuticals, including "Organic Fine Chemistry", "Biotechnology," and "Pharmaceuticals." With the exception of "Measurement" and "Furniture, Games," the effects are mostly positive but not significantly different from zero in the remaining fields. In particular, the effect for the field "Computer Technology" is large but imprecisely estimated. In Appendix 8.4 we obtain similar findings using different technology classifications.

The result that disclosure of technical information matters most for follow-on innovation in the fields of chemistry and pharmaceuticals is consistent with prior survey research, which reports that patents are effective in codifying knowledge in this fields and that inventors save time by reading such patents (Hall and Harhoff, 2012). According to Gambardella *et al.* (2011), median inventors save 27.59 hours by reading patents in "Organic fine chemistry" while they save only 1.35 hour in "Information Technology." Chemistry and pharmaceuticals are also fields where patents are particularly valuable and important for appropriability, as products in these areas are directly linked to specific molecular formations over which patent rights are relatively specific and clear(Cohen *et al.*, 2000).

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

A second analysis that can help inform our understanding of the mechanism by which libraries affect local patenting is one that focuses on the time period during which patent technical information

Figure 7: 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 25km of the library in year t and technological field τ. $PatentLib_i$ 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 2008 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 8.4 we report the results for alternative classifications of technological sub fields.

was more widely-available, i.e., the period after the introduction of Internet-based patent search engines, which began in the mid-1990s. A bit of background on the way in which patent technical information was available at patent deposit libraries is also helpful. Over time, the introduction of computers may have improved the usefulness of a patent library: Prior to 1982, most patent documents were available on microfilm but were not easily searchable. In 1982, the patent libraries for the first time introduced the possibility to search for prior art with a computer-based database. The search capabilities were "minimal in today's time frame, but a quantum leap in 1982" (Sneed, 1998). The new database, named "CASSIS," also enabled internet-like interconnection between libraries and the USPTO reducing communication costs and prior art search.

Beginning with the patent database of the USPTO in December 1994 and the Espacenet database of the EPO in December 1998, patent information became freely available on the internet. Around this time several commercial data vendors also started to sell databases. This might have reduced the utility of being close to a patent library. But even in 2004, 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." (Lambert, 2004) On December 14, 2006 Google Patents was launched starting the era of ubiquitous access to patent information.

To investigate how the effect of opening a patent library evolves over time we estimate Equation 1, using the following specification running the model separately for different set of library year openings:

$$\frac{\#Patents_{it}}{Population} = \sum_{\gamma \in \Gamma} [\beta_{1\gamma} \cdot Post_t + \beta_{2\gamma} \cdot PatLib_i \cdot Post_t] \cdot OpeningYearFE_{\gamma} + LibraryFE_i + YearFE_t + \varepsilon_i$$

where Γ is the partition of opening year in a set of five year periods. $Opening\,Year\,FE_{\gamma}$ is an indicator that is one if patent library i was opened in the five year period defined by γ. $\beta_{2\gamma}$ measures how much the patent libraries that were opened in the five year period defined by γ increases the average number of patents relative to their associated control libraries in the five years after the library was opened.

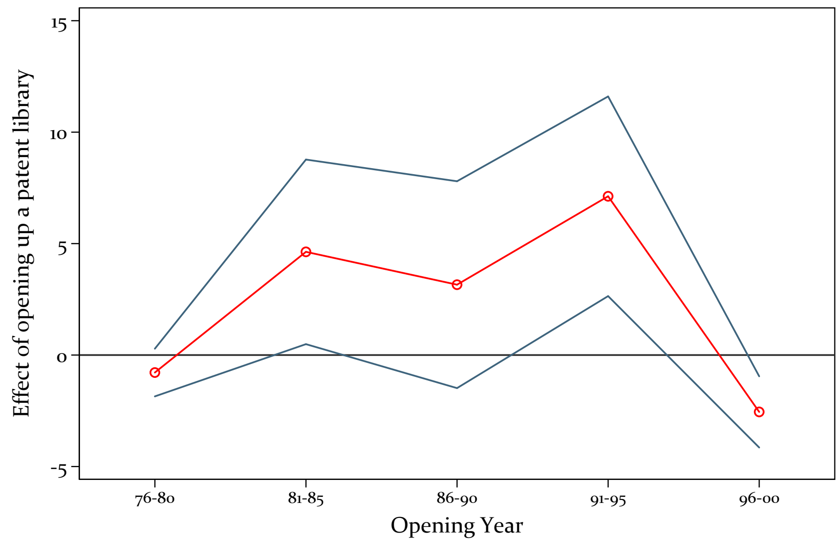
In Figure 15 we report the coefficients of $\beta_{2\gamma}$ over time. For the 1976 to 1980 period, the effect is indistinguishable from zero. During the periods 1981-1985, 1986-1990, and 1991-1995, the effect is large and stable. After reaching its maximum point in 1991-1995, a period during which the deposit library search capabilities had improved substantial, the effect declines beginning in 1996, the period corresponding to the widespread diffusion of the internet. These results speak in favor of a temporary increase in the importance of patent libraries due to new databases and a decline in importance when patent information became available on the internet. It is important to note that patent libraries continued to operate in the late 1990s as they had in the early 1990s, with the exception that the outside availability of patent information had changed. In Appendix 8.5 we interact the specification in Equation 1 instead with opening years with filing years and find a similar pattern.

While patent libraries opened the Internet era (post-1996) 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 8b 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.[18]

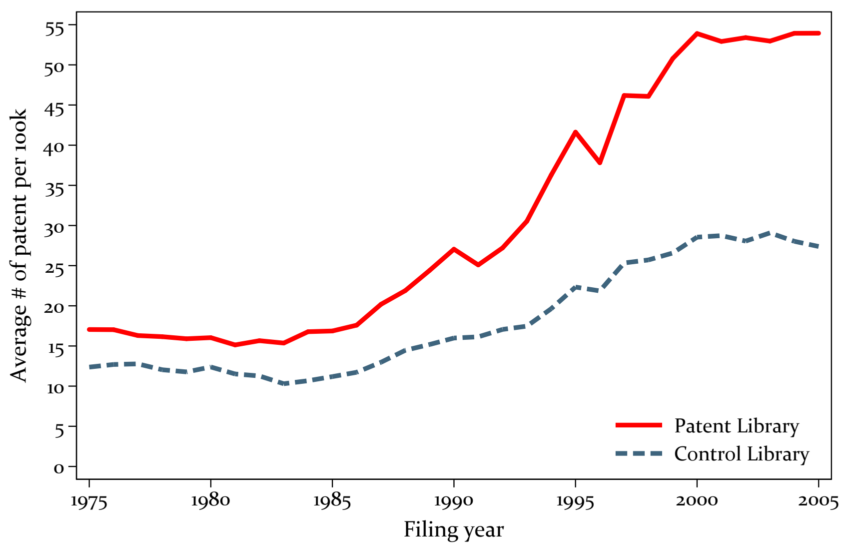
[18]Note, that this difference in patent numbers is (at best) the upper bound of the effect of the patent library program. The effect in our main regression is identified under the assumption that nothing else changes at the same time that increases patenting and is correlated with the opening of the patent library. 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 8: Effect over time

(a) Treatment effect by opening year period



(b) Averages over time



Note: Figure 8a 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 different five year periods of opening. To estimate this effect we use the following equation

$$\frac{\#Patents_{it}}{Population} = \sum_{\gamma \in \Gamma} [\beta_{1\gamma} \cdot Post_t + \beta_{2\gamma} \cdot PatLib_i \cdot Post_t] \cdot OpeningYear\ FE_\gamma + Library\ FE_i + Year\ FE_t + \varepsilon_i$$

and report $\beta_{2\gamma}$ for each filing year τ. We report 90% confidence intervals. Γ partitions all opening years in to a set of five year periods. Figure 8b shows the average number of patents around patent and control libraries per 100,000 persons over time. We classify all libraries as patent libraries that are opened from 1976 to 2001.

5.3 Structure of Patents: Citation Distance Increases, Quality of Patents Stays Constant

If the arrival of patent libraries in a region truly induce changes in innovation, such changes may be observable in differences 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 the average distance to cited patents increases after patent deposit libraries open. Likewise, if patent library opening simply induces patenting but not innovation, we would expect to that a higher fraction of post-opening patents in a region are of low quality and would, accordingly, receive fewer patent citations than in the pre-library period. To investigate these possibilities, we compare bibliometric features of patents associated with inventors in patent library regions with control patents of the same technology class and the same filing year but that are with 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 and once for the subset of inventors that patent before and after the opening of the patent library, i.e., "staying inventors." We cluster standard errors at the inventor level. As many patents only have a small number of backward citations, we condition on patents having five or more backward citations in analyses that include patent characteristics.[19] For these regressions we also condition on patents having the same number of backward citations.

Panel A of Table 5 reports the results of models assessing the impact of library opening on backwards citations make by patents in affected regions. Specifically, table reports the results of regressions of Equation 1, assessing changes in the amount and nature of backward citations, including backward citations make, the median of the geographic distances to the inventors of the cited patent, the originality of the citing patent, and the count of cited technology fields as dependent variables.[20] Column 1 shows that the average number of backward citations stays constant after patent library opening. The effect is the same for inventors that patented before the library opened (column 2). In columns 3 and 4 we use the log of the median geographic distance between citing and cited inventor as dependent variable. For all patents there is an increase in the median distance,

[19]In Appendix 8.7 we drop this requirement and the estimates become imprecise.

[20]To define originality of a patent we follow Hall *et al.* (2001)

Table 5: Impact of patent libraries on structure of patents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Backward Citations							
	All		> 5 Backward Citations					
	# Citations		log(Distance+1) Median		Originality		# Fields	
	All	Inv	All	Inv	All	Inv	All	Inv
Post x Pat lib	-0.3	-0.3	5.9*	4.9	1.4**	1.8**	2.4	3.4
	(0.2)	(0.5)	(3.6)	(5.9)	(0.6)	(0.8)	(3.3)	(4.8)
Obs.	622533	204402	112485	29398	112485	29398	112478	29395
	Panel B: Forward Citations							
	All		> 5 Backward Citations					
	# Citations		log(Distance+1) Median		Generality		# Fields	
	All	Inv	All	Inv	All	Inv	All	Inv
Post x Pat lib	0.4	1.9*	10.5**	4.2	1.3**	1.6*	8.5*	11.0
	(0.5)	(1.1)	(4.1)	(6.9)	(0.6)	(0.9)	(4.7)	(7.1)
Obs.	622533	204402	107336	28047	107336	28047	107257	28035

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:

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

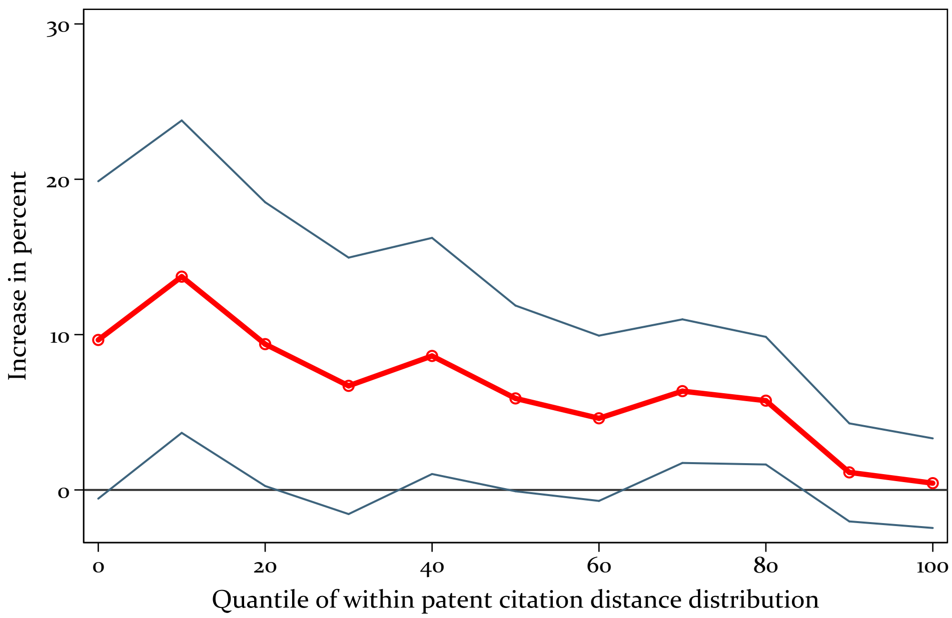
where $Outcome_{jt}$ is the outcome for the patent j that is filed in year t. $PatentLib_j$ is an indicator if patent j is around a patent library and $Post_t$ is an indicator for all years after the opening of the patent library. As controls we use a fixed effect for each combination of patent library, technology class and filing year. In column 1 of Panel A we use the sum of backward citation, in column 2 we use the median distance between the location of the inventor of the cited patent and the citing patent j. In column 5 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. The classification of technical fields follows Schmoch (2008). Then we take the average. In column 1 of Panel B we use forward citations, in column 2) the median distance to patents citing patent j, in column 3 the generality of patent j as defined by Hall *et al.* (2001) and in column 4 we count in how many technical fields the citing patents are filed. Standard errors are clustered on the inventor level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

consistent with the interpretation that access to prior art that was previously hard-to-access is driving the observed effects. The effect size corresponds to an increase of around 6% in distance across the set of patents in a region. Panel A of Figure 9 shows that citation distance increases most for those inventions that had low median distance before the introduction of libraries. This is consistent with what we would expect if patent access for previously-inhibited inventors were the driving mechanism behind the core findings - i.e., the set of patents previously citing the most geographically promixate sources experience the greatest boost in citing distance after libraries open. The results in columns (5) and (6) suggest that patents produced after patent library opening are more original. We the coefficients in (7) and (8) are consistent with the prospect that library patents cite a greater number of technological fields, these estimates are too noisy to be deemed statistically significant.

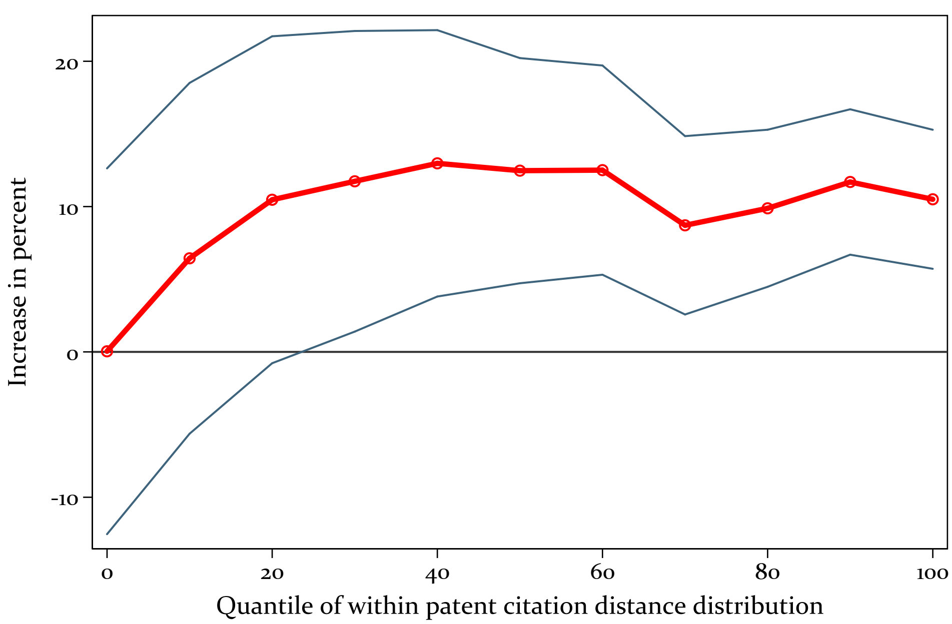
An additional important question regards whether the improved access to prior art changes the quality of the produced patents. This would be a particular concern if we were to find, for example, that post-library patents received substantially fewer citations than patents issued before library opening, as this would suggest a post-library decline in patent quality. To address these issues, we examine the characteristics of forward citations in library regions, including the number of forward citations, the distance of inventors citing the patent, the generality of the patent, and the number of different technical fields citing the patent. We report the results in Panel B of Table 5. The results do not suggest that induced patents are of lower quality than those produced before library opening. Columns (1) and (2) demonstrate that the average number of forward citations does not decrease after library opening. The coefficient in column (2) suggests that the number of forward citations actually increases for staying inventors, though to a degree that is only marginally statistically significant. Columns (3) and (4) document that the median distance increases between library region patents and the patents which cite them, although the effect is measured imprecisely for inventors that patented before. The effect size corresponds to an increase of around 10% in distance. Figure 9) Panel B, demonstrates that increasing distance of citations receive holds across nearly all deciles of pre-library citation distance - that is, for all but those deciles that had been cited only by the nearest patents, the average distance of citations received increases after patent library opening. This implies that library opening increases citing distance across the range of patents. Columns (5) and (6) document that patents become more general after library opening and columns

Figure 9: Effect by percentile of distance distribution

(a) Backward citations



(b) Forward citations



Note: This figure shows the increase in distance for each percentile of the within patent distance distribution. Subfigure a) shows the effects for the distance distribution of backward citations and Subfigure b) for the distance distribution of forward citations.

(8) and (9) document that they receive citations from an increasing number of technology fields.[21]

We interpret this set of results as demonstrating that patent library opening increases patenting as a result of providing access to previously-distant and expensive-to-access prior art along with increasing opportunities for subsequent innovation.

5.4 Census data: Opening of patent libraries is accompanied by "Creative Destruction"

To further explore whether patent libraries simply increase regional patenting or induce more meaningful changes in regional innovation output, we investigate business dynamics in library regions. To do this, we use Census data reporting business statistics for Metropolitan Statistical Areas (MSA) between 1977 and 2005.[22] These data report the number of firms, count of establishments, entry and exit of establishments, and job creation and job destruction at the MSA level. We focus on entry and exit of firms with fewer than 10 employees and on the subset of data that are not imputed for confidentiality reasons (Miranda and Vilhuber, 2014).[23] We scale each variable by population (in 100,000). As we do not observe the exact location for each company, we compare MSAs with a new patent libraries to MSAs with FDLs within the same state. While this approach involves a less precise comparison than our prior analyses, it still provides interesting insights into potential real industry effects of innovation via improved access to patents.

In Figure 10 we introduce descriptive statistics that will be consistent with our broader analysis of regional business dynamics in areas affected by patent libraries. Specifically, Figure 10 reports establishment entry and exit per capita in the years before and after an MSA receives a patent library. The figures show that the count of small establishments entries and exits increase significantly in MSAs that receive patent libraries relative those that do not. To test this more extensively, we estimate the impact of library opening on the battery of outcomes available in the 1977-2005 MSA-level Census data, using our standard differences-in-differences model. We plot the coefficient estimates for each outcome variable, along with its associated 95 percent confidence interval, in Figure 11. Panel A reports the results for establishments of firms with less than 10 employees and find library

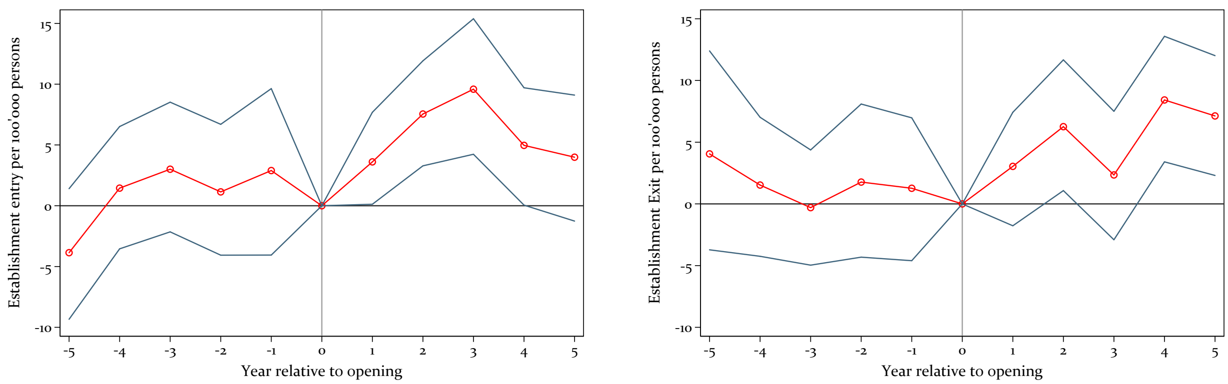
[21] To define the generality of a patent we follow Hall *et al.* (2001).

[22] The data is available on https://www.census.gov/ces/dataproducts/bds/data_estab.html

[23] In Section 8.8 in the appendix, we redo our analysis with the full dataset including the data that was censored and imputed for confidentiality reasons. We find very similar effects.

Figure 10: Business Dynamics

(a) Establishment Entry (b) Establishment Exit



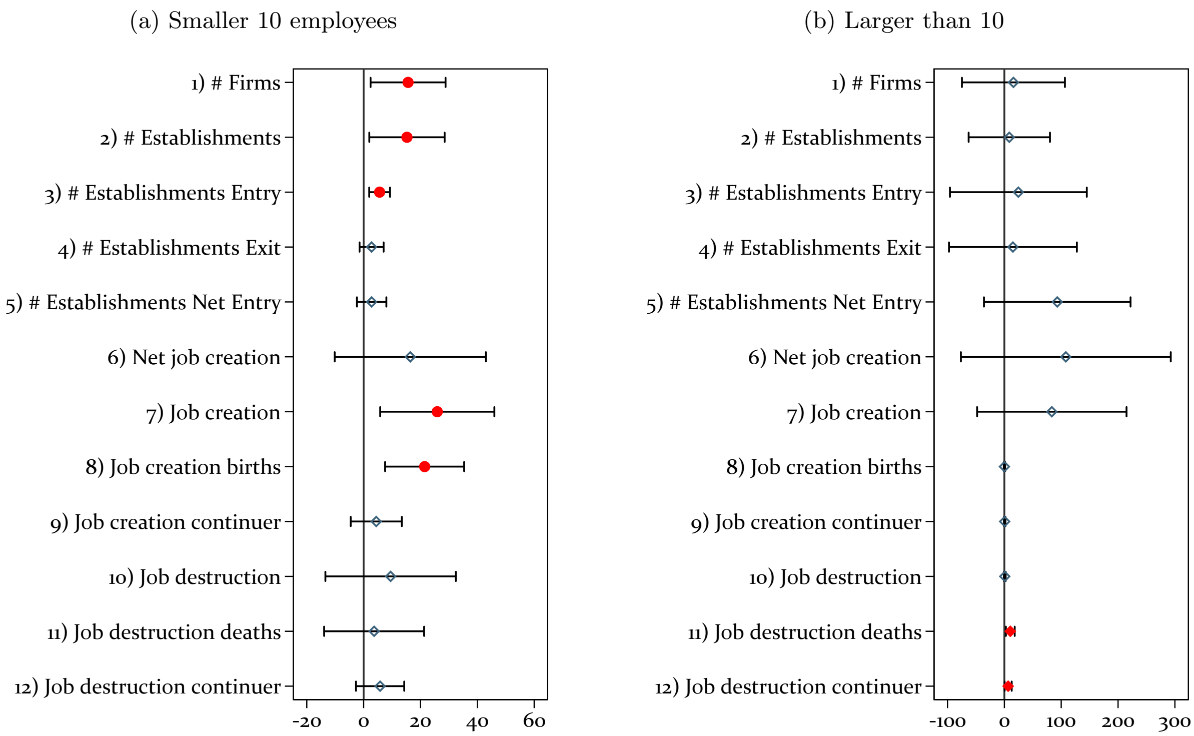
Note: This figure shows the yearly average treatment effects on the treated of opening up a patent library on entry (Panel a) and exit (Panel b) of establishment of companies with less than 10 employees in MSAs of patent libraries relative to MSAs of federal depository libraries in the same state. The 90% 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 Library within the same state, within 500 km but in different MSAs as control group. We exclude the patent libraries of Burlington and Rochester.

opening induces more firms, more establishments, more entry and job creation in particular by new companies.[24] But the effects for exit and job destruction (in particular 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 look at establishments of companies with more than 10 employees. We do not find any significant effect for entry and exit, but an effect for job destruction by firm deaths and continuing firms.

Taken together, these results point to an increase in the reallocation of workers and assets between firms. This is consisent with a process of creative destruction following increased innovative activity. Due to the opening of patent libraries, some entrants become innovative and grow at the expense of the not-so-innovative 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.

[24] To show the robustness of these estimates, in Section 8.9 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 11: 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 establihsments 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 partent library. $PatentLib_i$ 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.

6 Conclusion

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

Our study shows that the opening of patent libraries between 1977 and 1997 increased innovation around the libraries and helped to disperse innovative activity across the US. The results are largely driven by young and small companies and by patents in chemistry and pharmaceuticals. This innovation boost is not evident among regions that first received libraries in the period following the introduction of easily-accessible internet databases, although we find that the boost persists in this period among regions that received PDLs prior to the diffusion of the internet. The results also demonstrate that patent libraries were helping inventors to build on prior inventions by improving their access to technologically or spatially distant knowledge. We thus measure an "enablement effect" resulting from the disclosure of knowledge contained in patents. Finally, we also provide evidence that this increase in innovation led to job creation, particualrly by new small entrant establishments relative to larger incumbents.

Our estimates most likely provide a lower bound for the effect of patents on subsequent innovation through disclosure. First, in many public libraries, the titles and sometimes abstracts of patents were available in technical journals and books. Thus, even without a patent library, there might have been some awareness about inventions made elsewhere. Second, large companies often had their own patent library. Thus, we likely underestimate the effect of disclosure on the innovation of large businesses.

References

(2003). *OmniStats*. Tech. Rep. 4, US Department of Transportation.

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, forthcoming.

Biasi, B. and Moser, P. (2016). Effects of copyrights on science: Evidence from the wwii book republication program. *Available at SSRN 2542879*.

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.

— and Levinthal, D. A. (1989). Innovation and learning: The two faces of r&d. *The Economic Journal*, pp. 569–596.

—, 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*.

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.

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*.

Iacus, S. M., King, G. and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, **20** (1), 1–24.

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.

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*.

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

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. In *International Conference on Privacy in Statistical Databases*, Springer, 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.

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.

Petralia, S., Balland, P.-A. and Rigby, D. L. (2016). Unveiling the geography of historical patents in the united states from 1836 to 1975. *Scientific Data*, **3**, 160074.

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, 22 may 1997.

— (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.

Williams, H. L. (2017). How do patents affect research investments? *NBER Working Paper No. 23088*.

7 Appendix to Section2 & 3

Construction of Dataset

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 US patents.

2. We delete all foreign patents.

3. We geolocate all patents using the data of Li *et al.* (2014) for the publication years of 1975 to 2005 and of Balsmeier *et al.* (2017) for the years from 2005 to 2010. For geolocation of patents with a publication year before 1975 we use the data of Petralia *et al.* (2016)

4. We account for multiple counting of potential patents by author- and city-weighting patents. Here, we use all addresses of all inventors on a patent.

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 US.

6. We use population data from US Census on 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 6 and 7) is from Jenda (2005) and the complete list of Federal Depository Libraries is from the online Federal Depository Library Directory.

2. We drop the federal depository libraries in offshore areas: Pago Pago, AS; Mangilao, GU; Saint Thomas, VI; Kolonia, Pohnpei, FM; and Saint Croix, VI. We get the libaries' locations through their city and state information.

Table 6: List of all patent 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 (SC[I]3). Sunnyvale Public Library	1963
Madison, Wisconsin	Kurt F. Wendt Library. University of Wisconsin-Madison	1976
Birmingham, Alabama	Birmingham Public Library	1977
Dallas, Texas	Dallas Public Library	1977
Denver, Colorado	Denver Public Library	1977
Houston, Texas	Fondren Library. Rice University	1977
Raleigh, North Carolina	D.H. Hill Library. North Carolina State University	1977
Seattle, Washington	Engineering Library. University of Washington	1977
Lincoln, Nebraska	Engineering Library. University of Nebraska, Lincoln	1978
Sacramento, California	California State Library	1979
University Park, Pennsylvania	Schreyer Business Library. Paterno Library. Pennsylvania State Library	1979
Minneapolis, Minnesota	Minneapolis Public Library	1980
Newark, Delaware	University of Delaware Library	1980
Baton Rouge, Louisiana	Troy H. Middleton Library. Louisiana State University	1981
Albuquerque, New Mexico	Centennial Science and Engineering Library. The University of New Mexico	1983
Ann Arbor, Michigan	Media Union Library. The University of Michigan	1983
Auburn, Alabama	Ralph Brown Draughon Library. Auburn University	1983
Austin, Texas	McKinney Engineering Library. The University of Texas at Austin	1983
College Station, Texas	Sterling C. Evans Library. Texas A&M University	1983
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 7: 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

3. We geolocate patent libraries and Federal Depository Libraries using again patent data as all patent libraries are in places with at least one patent from 1975 to 2005. We assign all Federal Depository Libraries within 250 km to a patent library. If a Federal Depository Library would be assigned to multiple patent libraries we choose the closest one.

4. We drop all patent libraries that are not 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,. In our sample, 96% of patent depository libraries 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.

8 Appendix to Sections 3 & 4

8.1 Sample Characteristics and Results without Outliers

In Table 8, we showed descriptive statistics of the treatment and control observations in the year before the patent library openings. Some observations do not show patenting activity in some years. In Table 8, we delete these outliers. This results in a more balanced sample.

In Table 9 we show that when we use this sample for our main estimation, our results are qualitatively identical and even slightly larger.

8.2 Compare Averages

In Figure 12a we compare the raw difference in the average numerb of patents per 100'000 persons around treatment and control libraries. In Figure 12b we substract 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 diverge in after the patent library opened.

Table 8: Summary statistics in the year before opening, deleting outlier observations

	Patent Libraries	Control Libraries	Diff	P-Value
# Patents/100k	17.79	14.41	-3.39	0.19
# Patents	132.73	133.55	0.82	0.97
# Pat. small firms/100k	8.27	6.02	-2.25	0.08
# Pat. big firms/100k	9.52	8.39	-1.14	0.53
# Pat. young firms/100k	0.16	0.18	0.02	0.81
# Patents old firms/100k	17.63	14.23	-3.40	0.18
Population in 100k	7.90	14.52	6.62	0.00
Uni Library	0.71	0.74	0.02	0.71

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

Table 9: Patent libraries and local innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Baseline		Size		Age		University	
	250km	Citations	Dollars	Small	Large	Young	Old	Yes	No
Post	-0.6	-18.4	-4.6	-0.5	-0.1	2.8***	-3.4***	-1.0	-1.6*
	(0.7)	(22.0)	(3.1)	(0.4)	(0.4)	(0.3)	(0.6)	(0.8)	(0.9)
Pat Lib x Post	3.3**	62.6***	10.7***	1.4*	1.9	2.1**	1.2	4.3**	0.4
	(1.3)	(19.6)	(3.9)	(0.7)	(1.2)	(1.0)	(1.1)	(1.7)	(1.5)
Mean Dep.	16.5	230.2	61.1	7.6	8.9	3.3	13.3	15.8	18.6
R2 (within)	0.12	0.12	0.21	0.19	0.05	0.48	0.10	0.10	0.30
Obs.	4037	4037	4037	4037	4037	4037	4037	2981	1056

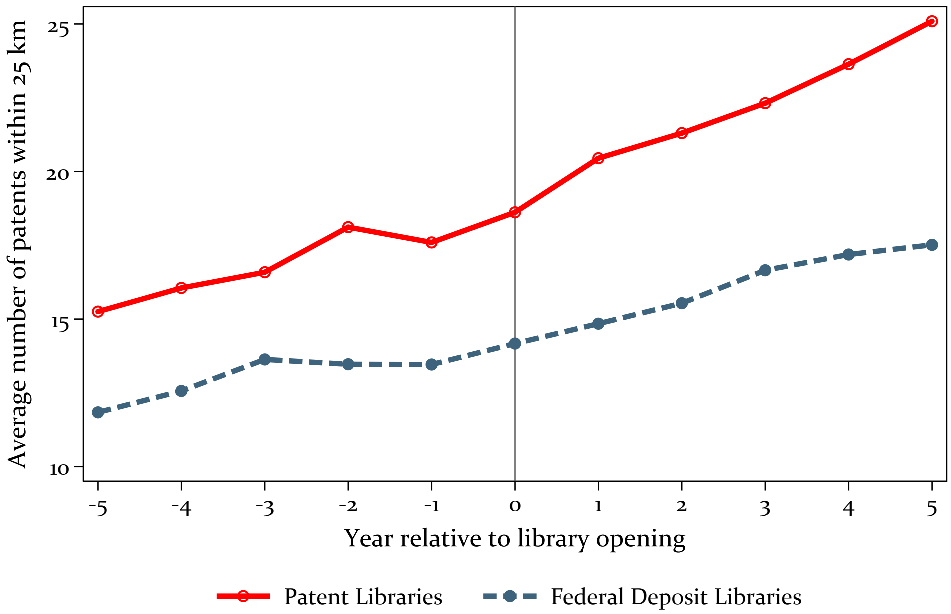
Note: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is:

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

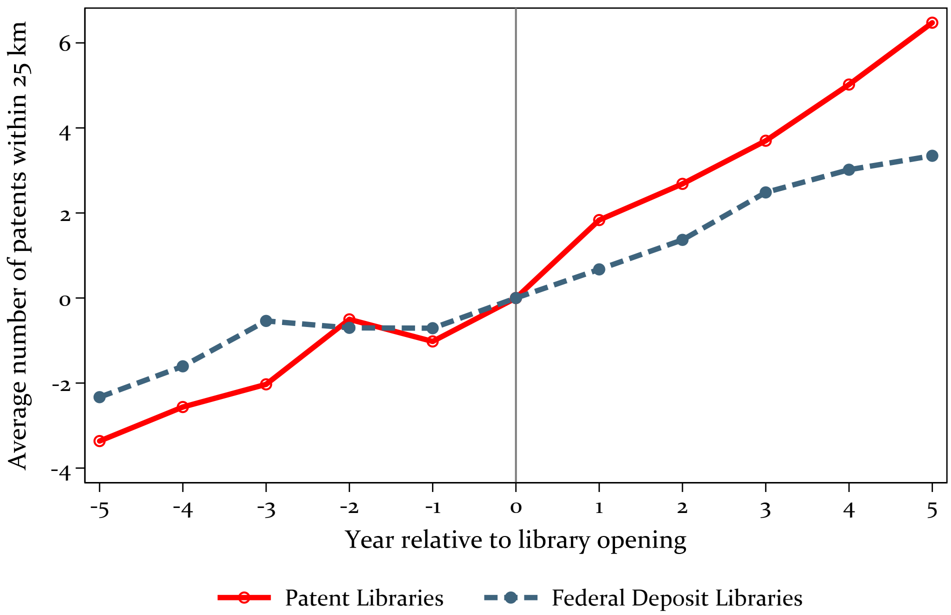
where $PatentLib_i$ is an indicator if the library i is a patent library and $Post_t$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within 250km as controls. In comparison to our main estimation results, we delete outlier libraries that report years without any patenting in the region. In column (2) we weigh each patent with its forward citations. In column (3) we use the patent values from Kogan *et al.* (2017) to weigh each patent with its dollar value. We windsorize the Dollar values at the 90th percentile to adjust for outliers. In column (4) and (5) 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 the following two columns we split the dependent variable by young and old assignees. An assignee is young if it filed its first patent after the opening of the library and old otherwise. 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.

Figure 12: Compare Averages

(a) Raw



(b) Normalized in the opening year

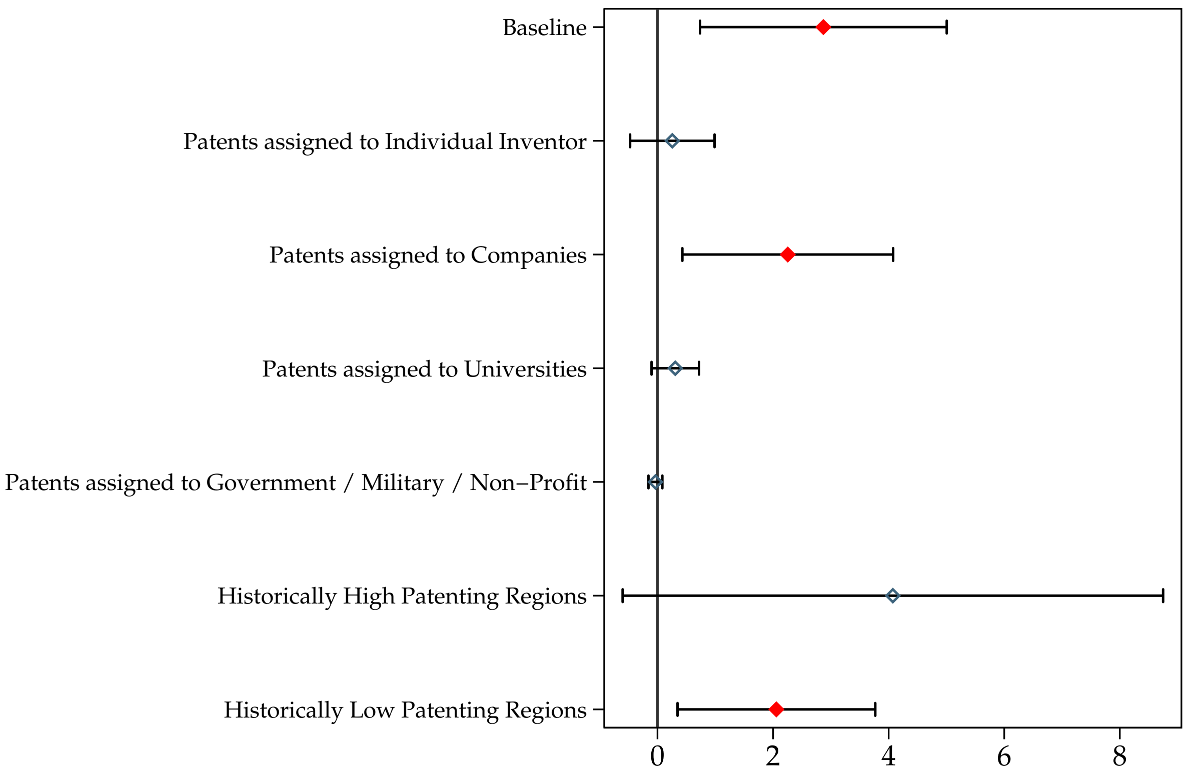


Note: This figure plots the average number of patents within 25 km of the patent library (red solid line) and around Federal Depository Libraries (blue dashed line) in the five years before and after the opening of the library. Subfigure 12a shows the raw average and in Subfigure 12b we normalize the average relative to its value in the year of the opening.

8.3 Further Sample Splits and Robustness

In Figure 13 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. Then we split the sample in historically high and low patent regions. The effec is similar in both, but a bit stronger in historically low patenting regions. In the last two lines we use the information if the patent library is also a university library to split the sample. We find that the effect is positive for both but stronger for patent libraries that are also university libraries.

Figure 13: 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:

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

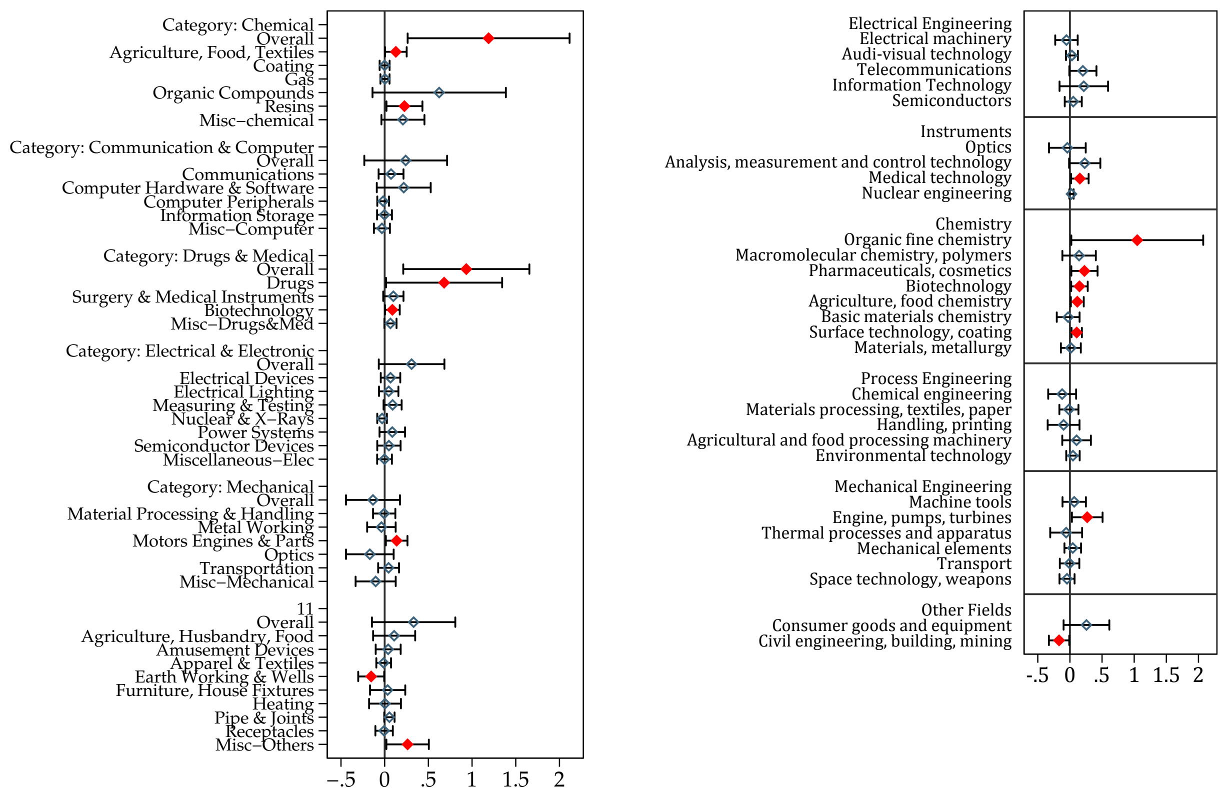
where $PatentLib_i$ is an indicator if the library i is a patent library and $Post_t$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In the first line we report the point estimate for $\beta_2$along with 90% confidence intervals. The confidence intervals are based on standard errors that are clustered on the patent library level. In line 2 to 5 we split the dependent variable by the type of assignee. We show results seperately for independent inventors, patents assigned to companies, patents assigned to universities and patents assigned to the government, military or non-profits. In line 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.

8.4 Alternative Technology Classifications

In Figure 14 we use two alternative technology classification to show the effects across fields. In Subfigure 14a we use the NBER subcategory that are based on the USPTO technology classes. In Subfigure 14b 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 14: Alternative technology classifications

(a) By NBER subcategory (USPTO tech classes based)   (b) ISI-OST-INPI Technological Categories 1995



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 25km of the library in year t and technological field τ. $PatentLib_i$ 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 1995 if 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 8.4 we report the results for alternative classifications of technological subfields.

8.5 Alternative Estimation of the Effects over Time

In the results section we estimate the effect of opening a patent library separately for different opening years. In this section we estimate the effect for each filing year. We use the following equation to arrive at the effect over time:

$$\frac{\#Patents_{it}}{Population} = \sum_{\tau=1976}^{1999} [\beta_{1\tau} \cdot Post_t + \beta_{2\tau} \cdot PatLib_i \cdot Post_t] \cdot Year\,FE_t + Library\,FE_i + Year\,FE_t + \varepsilon_i$$

where $Year\ FE_t$ is an indicator that is one if τ is equal to the filing year t. In addition we transform the regression from time relative to opening to filing year time. We look at the five years before and after the opening of the patent library. β_{2t} measures how much the patent libraries that were opened in $t-5$ to t increases the average number of patents relative to their associated control libraries in year t.[25]

In Figure 15 we report the coefficients of β_{2t}over time. From 1976 to 1985, the effect is positive and close to the average effect we find for the whole sample. In 1985 the effect increases in 1986 and declines from 1995 onwards.[26] These results speak in favor of a temporary increase in the importance of patent libraries due to new databases and a decrease in importance when patent information became available over the internet. In Appendix 8.5 we interact 1 instead of filing years with opening years and find a similar pattern.

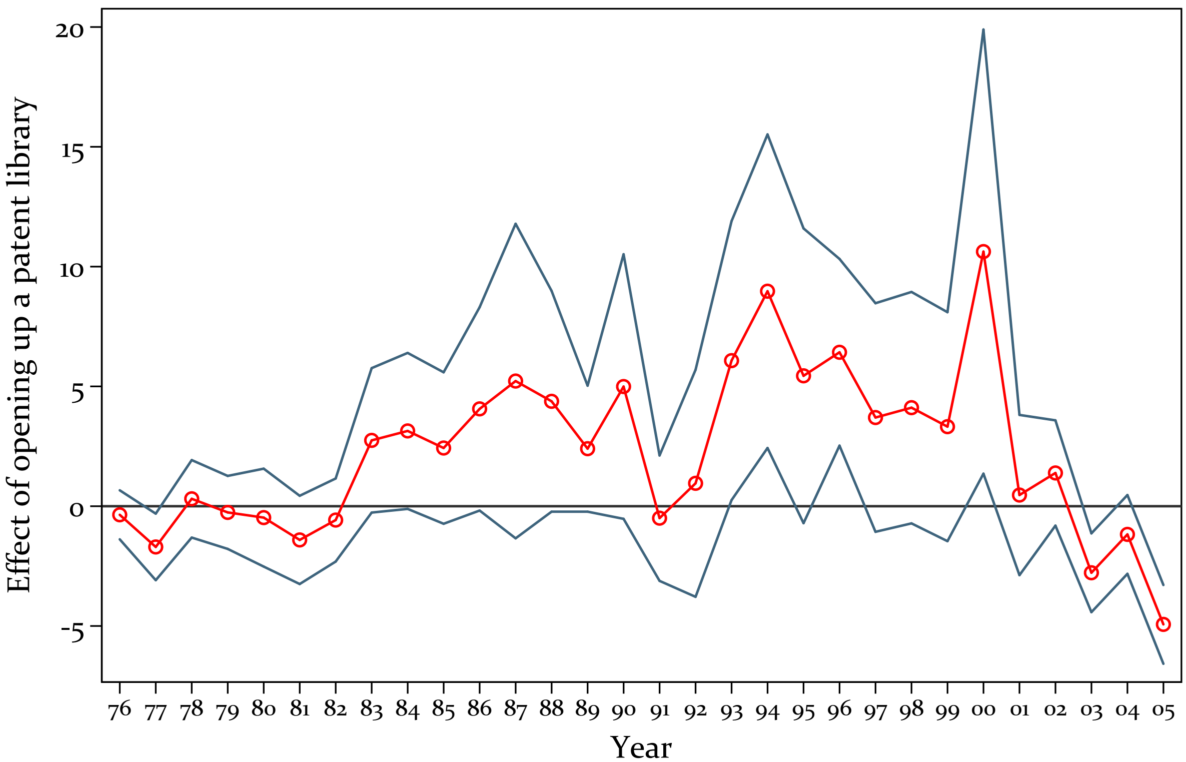
8.6 Synthetic Development of Patenting if Share had Remained Constant

To use an alternative benchmark to quantify the impacts of patent libraries, we assess how much patents would have been expected in library regions if patenting relative to other regions had remained constant. That is, we fix the share of patents around libraries among all US patents to its average in the five years before the library opened. We then compare the actual to this counterfactual number of patents, on a per-capita-level. As can be seen in 16, the results resemble earlier conclusions: Relative to what would have been expected under constant shares among all patents, regions around patent libraries clearly increase their patenting.

[25]The interpretation of these yearly coefficient of this regression is different from the standard difference-in-difference set-up. β_{1t} identifies the difference between the group of patent libraries (and their associated control libraries) that in year t are already opened relative to those that are not. β_{2t} identifies within for the set of opened patent librariers the difference in number of patents to their control libraries.

[26]The estimation ends at 1999 as in 2000 the last patent library opens and thus we cannot identify β_{1t} any more.

Figure 15: Effect of opening a patent library over time

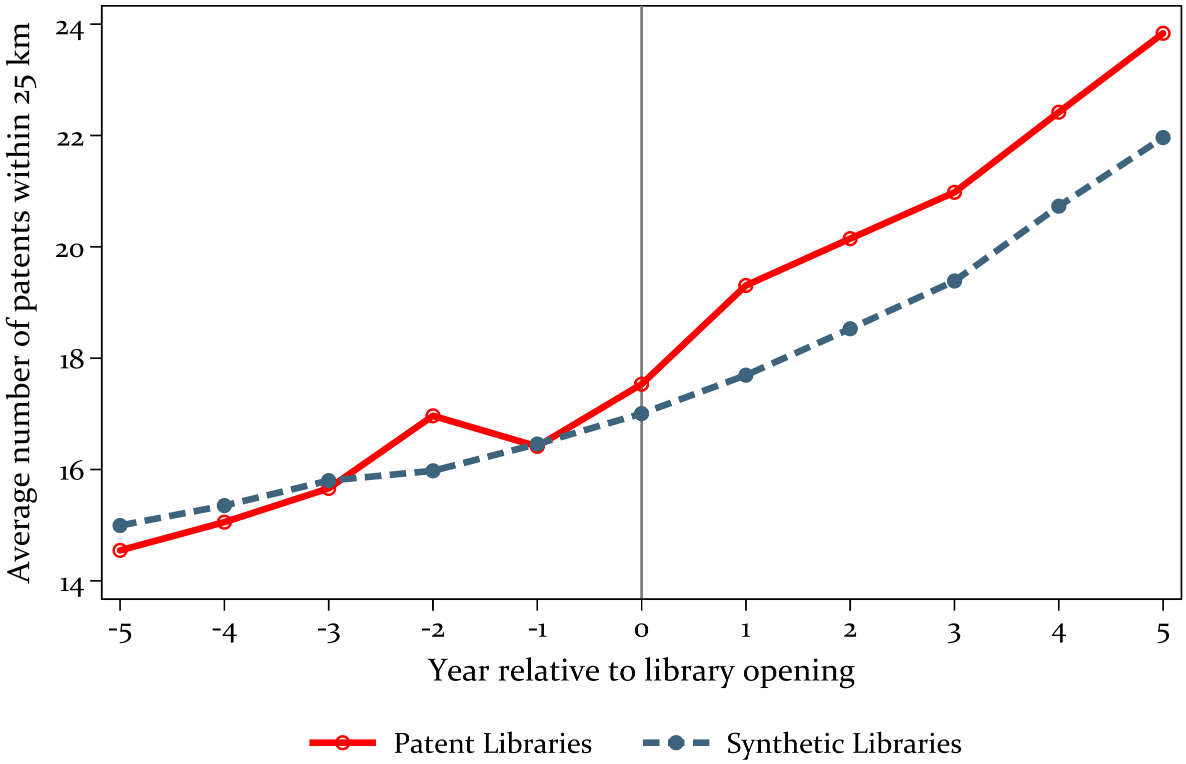


Note: This figure shows the average treatment effects on the treated of opening up a patent library on the average number of patents in a particular filing year within 25 km of patent libraries relative to the average number of patents around matched federal depository libraries. We assign each patent library and all Federal Depository Library within the same state and within 250 km as control group and use the ten years before and after the library opened as estimation sample. To estimate this effect we use the following equation

$$\frac{\#Patents_{i\tau}}{Population} = \sum_{t=1976}^{1999} [\beta_{1t} \cdot Post_t + \beta_{2t} \cdot PatLib_i \cdot Post_t] \cdot Year\,FE_t + Library\,FE_i + Year\,FE_t + \varepsilon_i$$

and report β_{2t} for each filing year τ. We report 90% confidence intervals.

Figure 16: Synthetic libraries: constant share among U.S. patents



Note: This figure shows the actual number of patents per capita (in red) relative to the counterfactual number of patents per capita had the share of patents around PTDLs among all U.S. patents remained constant (blue). To arrive at these, we use the average share of patents around PTDLs in the five years before the library opening.

8.7 Structure of patents for the full sample

In Section 5.3 we condition on a patent having at least five backward citations to aid measurement. In Table 10 we drop this requirement and all the estimates are mostly in the same direction but much less precise.

8.8 Census data with imputed data

In Section 5.4 we delete data that is marked as imputed for confidentiality reason (d_flag==1). As this is certainly a non-random sample we repeat here our analysis with all, imputed and non-imputed, observations. The results are very similar with the exception that job creation of births is not significantly positive anymore.

Table 10: Impact of patent libraries on structure of patents - Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Backward Citations							
	All		All					
	# Citations		log(Distance) Mean		Originality		# Fields	
	All	Inv	All	Inv	All	Inv	All	Inv
Post x Pat lib	-0.3	-0.3	-3.4	-19.9	0.6	0.9	0.0	0.0
	(0.2)	(0.5)	(17.6)	(31.7)	(0.4)	(0.6)	(0.0)	(0.0)
Obs.	622533	204402	245934	63341	245934	63341	244925	63092

	Panel B: Forward Citations							
	All		All					
	# Citations		log(Distance) Mean		Generality		# Fields	
	All	Inv	All	Inv	All	Inv	All	Inv
Post x Pat lib	0.4	1.9*	11.1	11.5	0.6	0.4	0.0	0.0
	(0.5)	(1.1)	(19.8)	(37.5)	(0.5)	(0.8)	(0.0)	(0.1)
Obs.	622533	204402	229284	58925	229284	58925	228984	58859

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:

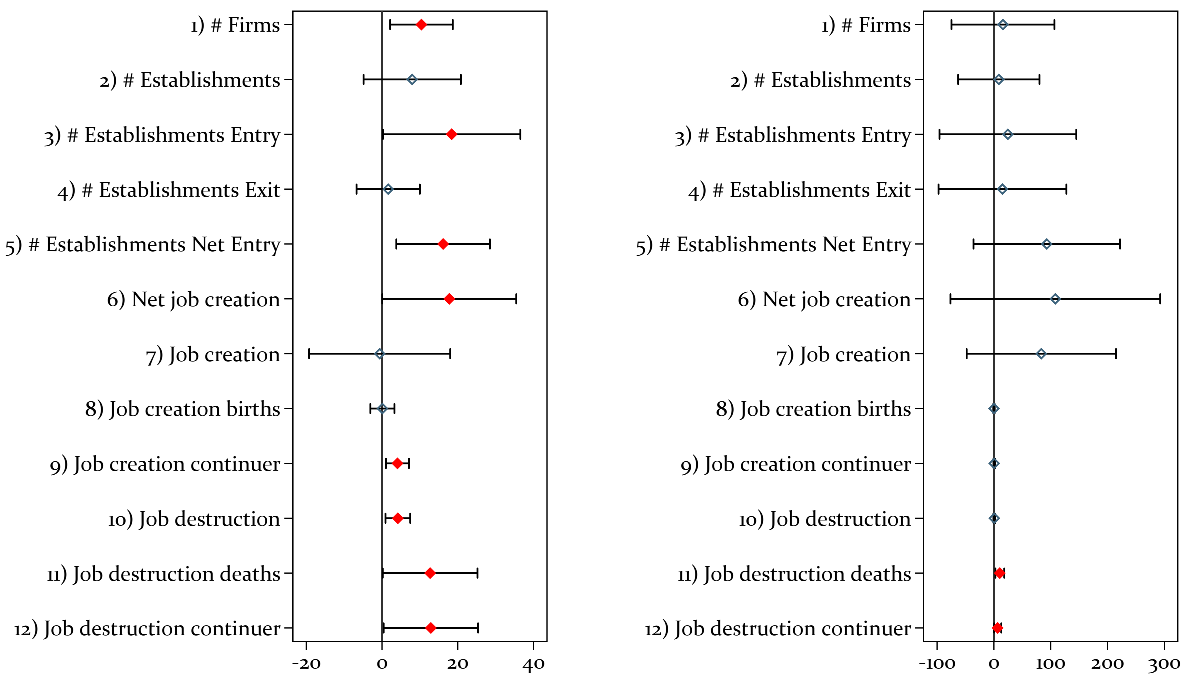
$$Outcome_{jt} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_j \cdot Post_t + Controls + \varepsilon_i$$

where $Outcome_{jt}$ is the outcome for the patent j that is filed in year t. $PatentLib_j$ is an indicator if patent j is areound 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 column 1 of Panel A we use the sum of backward citation, in column 2 we use the average 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 calcluate for each citation link the likelihood that a patent in the patent class of j cites a patent in the patent class of the cited patent. Then we take the average. In column 1 of Panel B we use forward citations, in column 2) the average distance to patents citing patent j, in column 3 the generality of patent j as defined by Hall *et al.* (2001) and in column 4 the average likelihood that a patent in the patent class of j is cited by patent in the patent class of the citing patents. Standard errors are clustered on the inventor level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure 17: Difference in Difference: Census Data

(a) Smaller 10 employees

(b) Larger than 10

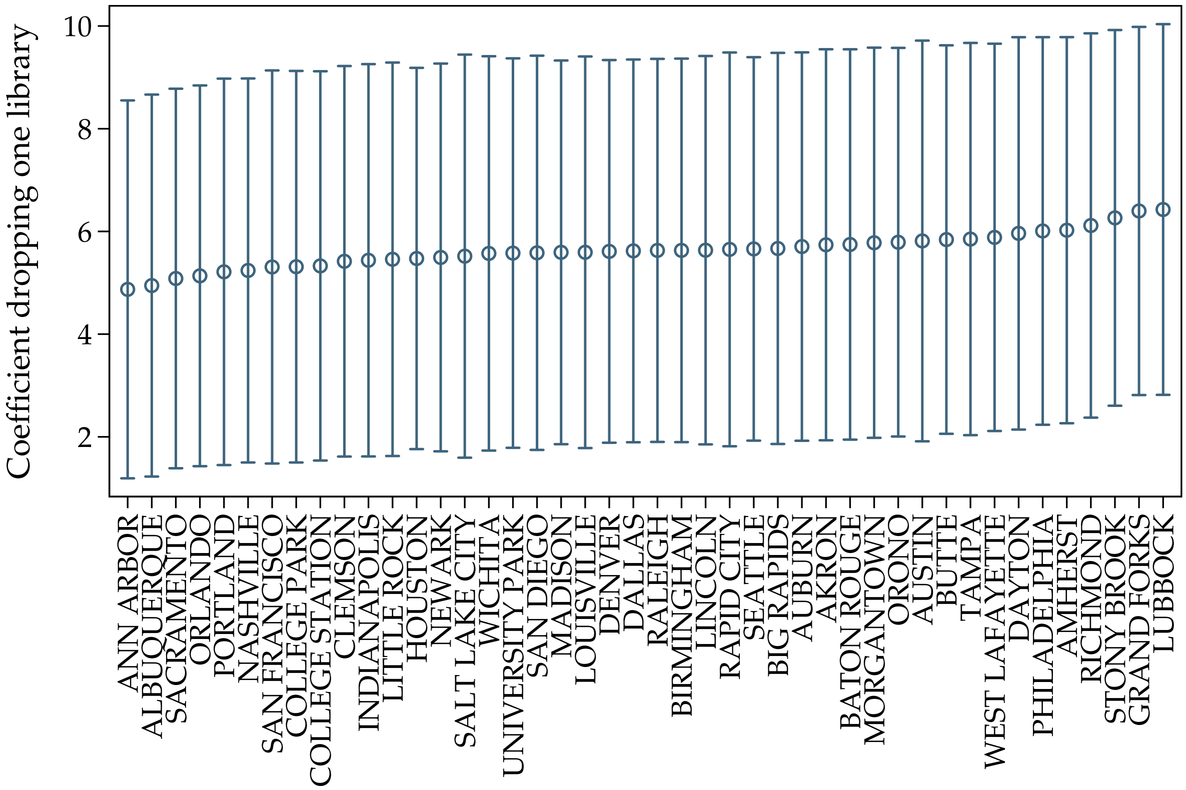


Note: This figure shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. In Subfigure a) the sample are all establishments smaller than 10 employees and in Subfigure b) all establihsments larger than 10. 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 partent library. $PatentLib_i$ 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.

Figure 18: Stability: Leave-one library out estimator



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 partent library. $PatentLib_i$ 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.

8.9 Census results: robustness

To show that the real industry effects do not depend on any patent library in particular, we repeat our leave-one-out analysis for our main result on business dynamics, namely the entry of new establishments. Figure 18 shows results when we drop libraries one by one. The coefficient does not change qualitatively, showing the robustness of our estimates.