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

Furman, Nagler, Watzinger

Online Appendix

Appendix to Sections 2 & 3

A1. List of All Opened Patent Libraries

Table A1 and Table A2 show a list of all Patent Depository Libraries in our data, following Jenda (2005).

Table A1—: List of all Patent Depository Libraries

City, State	Name of Library	Opening Year
Albany, New York	New York State Library Cultural Education Center	1870
Boston, Massachusetts	Boston Public Library	1870
Columbus, Ohio	Science and Engineering Library. Ohio State University	1870
Los Angeles, California	Los Angeles Public Library	1870
New York, New York	New York Public Library	1870
St. Louis, Missouri	St. Louis Public Library	1870
Buffalo, New York	Buffalo and Erie County Public Library	1871
Cincinnati, Ohio	The Public Library of Cincinnati and Hamilton County	1871
Detroit, Michigan	Great Lakes Patent and Trademark Center. Detroit Public Library	1871
Chicago, Illinois	Chicago Public Library	1876
Newark, New Jersey	Newark Public Library	1880
Cleveland, Ohio	Cleveland Public Library	1890
Providence, Rhode Island	Providence Public Library	1901
Pittsburgh, Pennsylvania	The Carnegie Library of Pittsburgh	1902
Toledo, Ohio	Toledo/Lucas County Public Library	1934
Atlanta, Georgia	Library and Information Center. Georgia Institute of Technology	1946
Kansas City, Missouri	Linda Hall Library	1946
Milwaukee, Wisconsin	Milwaukee Public Library	1949
Stillwater, Oklahoma	Patent and Trademark Library. Oklahoma State University	1956
Sunnyvale, California	Sunnyvale Center for Innovation, Invention & Ideas, Sunnyvale Public Library	1963
Madison, Wisconsin	Kurt F. Wendt Library. University of Wisconsin-Madison	1976
Birmingham, Alabama	Birmingham Public Library	1977
Dallas, Texas	Dallas Public Library	1977
Denver, Colorado	Denver Public Library	1977
Houston, Texas	Fondren Library. Rice University	1977
Raleigh, North Carolina	D.H. Hill Library. North Carolina State University	1977
Seattle, Washington	Engineering Library. University of Washington	1977
Lincoln, Nebraska	Engineering Library. University of Nebraska, Lincoln	1978
Sacramento, California	California State Library	1979
University Park, Pennsylvania	Schreyer Business Library. Paterno Library. Pennsylvania State Library	1979
Minneapolis, Minnesota	Minneapolis Public Library	1980
Newark, Delaware	University of Delaware Library	1980
Baton Rouge, Louisiana	Troy H. Middleton Library. Louisiana State University	1981
Albuquerque, New Mexico	Centennial Science and Engineering Library. The University of New Mexico	1983
Ann Arbor, Michigan	Media Union Library. The University of Michigan	1983
Auburn, Alabama	Ralph Brown Draughon Library. Auburn University	1983
Austin, Texas	McKinney Engineering Library. The University of Texas at Austin	1983
College Station, Texas	Sterling C. Evans Library. Texas A&M University	1983
Indianapolis, Indiana	Indianapolis-Marion County Public Library	1983
Moscow, Idaho	University of Idaho Library	1983
Reno, Nevada	University Library. University of Nevada-Reno	1983
Amherst, Massachusetts	Physical Sciences and Engineering Library. University of Massachusetts	1984
Anchorage, Alaska	Z. J. Loussac Public Library. Anchorage Municipal Libraries	1984
Butte, Montana	Montana Tech Library of the University of Montana	1984
College Park, Maryland	Engineering and Physical Sciences Library. University of Maryland	1984
Fort Lauderdale, Florida	Broward County Main Library	1984
Miami, Florida	Miami-Dade Public Library System	1984
Salt Lake City, Utah	Marriott Library. University of Utah	1984
San Diego, California	San Diego Public Library	1984
Springfield, Illinois	Illinois State Library	1984

Table A2—: 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

A2. Dataset Construction

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

Patent Data

1) We use patent data from the PATSTAT Database of the EPO (European Patent Office, 2016) that contains the universe of U.S. patents.
2) We delete all patents that pertain to foreign inventors.
3) We geolocate all patents using the data of Balsmeier et al. (2018) and Morrison, Riccaboni and Pammolli (2017).
4) We account for patents with inventors in multiple cities by using city-weighted patents.
5) To calculate citation distance, we assign the address of the first inventor on the citing or cited patent to the entire patent. When there is no primary inventor, we keep the first one in the list. We use only citations that are within the U.S.
6) We use population data from 2010 U.S. Census at the level of the incorporated city and compute yearly patent and citation rates per capita in circles around all library locations.

Library Data

1) Data on patent libraries (see tables A1 and A2) are 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 outside the continental United States, including Pago Pago AS; Mangilao GU; Saint Thomas VI; Kolonia, Pohnpei FM; and Saint Croix VI. We obtain the library location information based on their city and state.
3) We geolocate patent libraries and Federal Depository Libraries using patent data, as all patent libraries are in places with at least one patent between 1975 to 2005. We match all Federal Depository Libraries within 250 miles to a patent library. If a Federal Depository Library can be assigned to multiple patent libraries, we match it to the geographically closest patent library unless both patent libraries are almost equally close (less than 5 miles difference in distance).

4) We drop all patent libraries that are not also Federal Depository Libraries at the time of patent library opening. To obtain a better match of treatment and control library we delete all small federal depository libraries because patent depository libraries are usually either medium sized or large federal depository libraries. Of the patent libraries that were opened in our sample period that are also FDLs, 96% are considered medium sized or large, and only three patent libraries are considered small.

In a last step we cross all inventor locations with our library data to obtain pair-wise combinations of locations between inventors and patent libraries.

Patent text

1) We combine three sources of patent text data:
   a) All titles and abstracts of patents from Patstat for patents in English before 1985 and all US patents from 1985 to 2013.
   b) The text of all patent claims of all US patents from 1975 to 2013 from the PatentsView (PatentsView, 2020) database (https://www.patentsview.org/download/).
   c) The unique set of words in the abstract and the title of all US patents from 1975 to 2013 from Arts, Cassiman and Gomez (2018).
2) We partition the set of patent in the following subsets: A patent with
   a) "Words already used in the region": All words in the patent were used before the library opening in any published patent. We define all words used in any patent before 1970 as already used in all regions to ameliorate the problem that words that are old but only rarely used might be otherwise classified as new.
   b) "New to region but not world": The patent includes at least one word that is used for the first time in the region but that was used in the other published patens before.
   c) "New to the world": The patent includes at least one word that is used for the first time in any published patent.
   d) "Words that appeared in region after opening": The patent includes at least one word that was used for the first time after the library opening.
3) If a patent is classified in several of the categories above we give priority to "New to region but not world" over any other category and "New to the world" over "Words that appeared in region after opening". The reason is that we are mostly interested in knowledge transfer and "New to region but not world" is the best measure for knowledge transfer.

   Note that because the patent text (see 1b and 1c) is only available starting in 1975, we need to restrict our analysis to patents filed after this date.

Appendix to Sections 3 & 4

B1. Summary Statistics without Outlier Regions

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

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

Main sample	Patent Libraries	Control Libraries	Diff	P-Value
Population in 100k	7.25	4.70	-2.56	0.13
Uni Library	0.69	0.70	0.02	0.87
# Patents	135.13	76.99	-58.14	0.10
# Patents/100k	17.71	14.52	-3.19	0.32
Citation-weighted patents	259.45	207.90	-51.55	0.37
# Pat. small firms/100k	7.94	7.28	-0.65	0.58
# Pat. big firms/100k	9.77	7.24	-2.53	0.30
# Pat. young firms/100k	5.71	5.18	-0.53	0.54
# Patents old firms/100k	12.00	9.34	-2.65	0.32
Number of libraries	45	256		

Patents by field	Patent Libraries	Control Libraries	Diff	P-Value
Electrical Engineering	2.67	2.44	-0.23	0.75
Instruments	2.74	2.21	-0.53	0.32
Chemistry	4.82	2.75	-2.06	0.18
Process Engineering	2.53	2.72	0.19	0.74
Mechanical Engineering	2.82	2.46	-0.36	0.54
Other Fields	2.11	1.92	-0.19	0.58

Note: This table shows the averages of the data for patent libraries and control libraries without outlier regions that report zero patenting in the five years before or after patent libary opening. The last two columns shows differences with the associated significance levels. A firm is defined as young if its first patent was filed less than three years before the opening of the patent library, otherwise it is old. A firm is defined as small if it has no more than 20 patents before the opening of the patent library, otherwise it is large. The p-values result from a t-test with unequal variances.

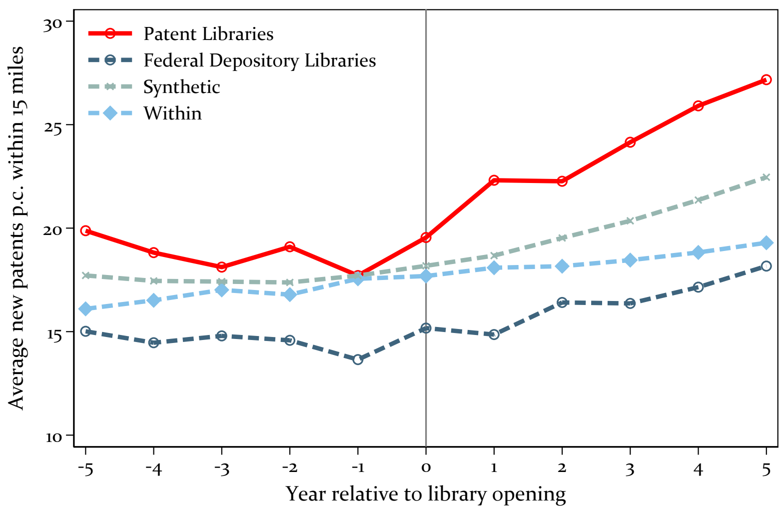
B2. Compare Averages

In the main part of the paper we employ three fundamentally different approaches to constructing the control group for the patent libraries. In Figure B-1a we report the raw difference in the average number of patents per 100,000 persons around treatment and the three control groups. As a first control group, we use Federal Depository Libraries within 250 miles around the patent library. As a second control group, we construct a "synthetic" doppelgaenger for each patent library. We do this by calculating how many patents would be around a patent library in the years after the opening if the average share of patents of the region among all patents in the U.S. would have remained constant to the year before opening. The third control group involves patent libraries that are opened at a future date ("within"). In this Figure, we use all future patent libraries as controls of the library to keep the sample constant. In the body of the paper , we drop patent libraries from the sample as soon as they are opened.

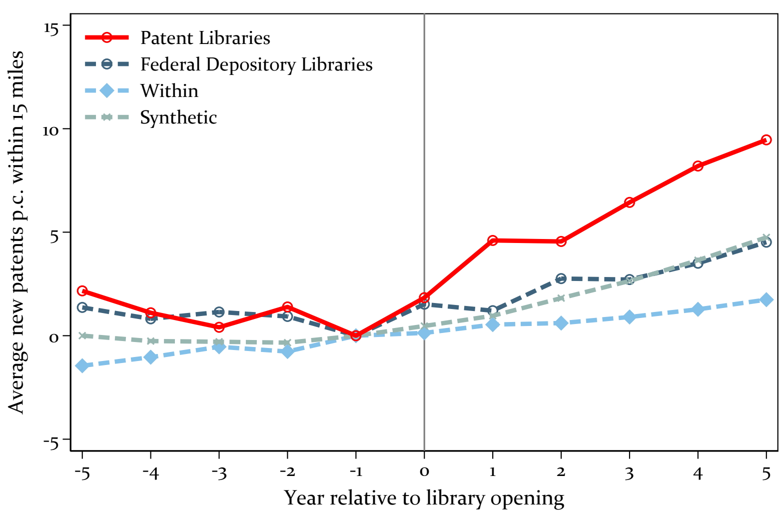
Before patent library opening, the number of patents around treatment and control are stable. After the opening of a patent library, there are more patents filed around the patent libraries. There is also an upward trend among all potential control groups. To better see the relative increase, we subtract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

Figure B-1. : Compare Averages

(a) Raw



(b) Normalized in t=-1

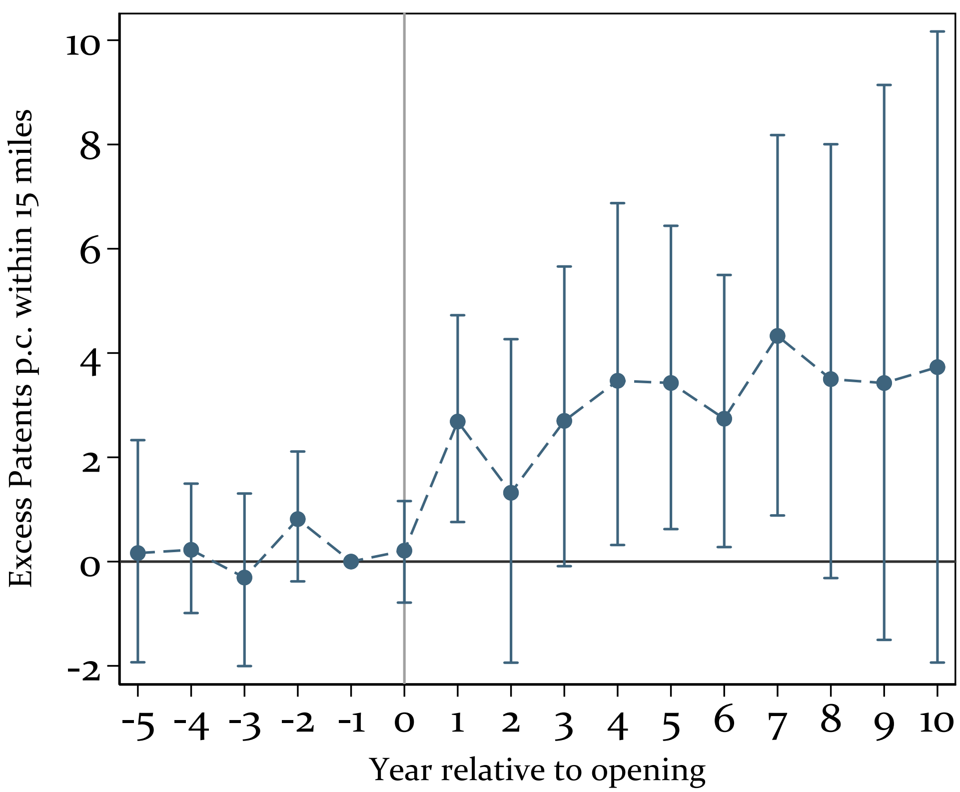


Note: This figure plots the average number of patents within 15 miles of the patent library (red solid line), around Federal Depository Libraries (dark blue dashed line), around synthetic patent library regions (green dashed line), and around patent libraries that are opened later (light blue dahed line)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

B3. Time-varying Treatment Effects Using a Longer Time Window

To assess how the estimates develop after our main five-year window after patent library opening, we repeat our estimation using a longer time window. Figure B-2 shows the results. The effects seem to even increase in the longer-run. We are however reluctant to speculate on the longer-term effects of Patent Deposit Libraries. Expanding the time window makes it much harder to assess potential mechanisms behind these results as many things may have changed over time. Thus, our confidence in the exogeneity of PDL opening declines as we expand the window for analysis.

Figure B-2. : Main treatment effect, longer time window



B4. Further Sample Splits and Robustness

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

Table B-2 shows the results underlying Figure 6.

In Table B-3, we provide additional robustness tests for our main specification. In the first column we show our baseline estimate. In the second column, we repeat the estimation without using the weights of the CEM algorithm. In the third column, we include patent library-specific trends in the regression. In the fourth column, as the dependent variable we only use the per capita count of patents whose inventors are within 15 miles of each other or patents from solo inventors. In the final two columns, we show the robustness of our results when using larger circles around patent libraries to compute patents per capita with 25 and 50 miles, respectively. Finally, in the last two columns we use the number and the log number of patents in the 15 miles around the libary as the respective dependent variable, controlling for (now time-varying) population. Our results are robust to all of these robustness tests.

Table B-2—: Auxiliary Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
		Add matching on			Distance Match				Patents p.c. in		
	Baseline	University	University + Patents p.c.	University + Patents p.c. + Population	¡ 100 mi	5 closest	Synthetic libraries	Pseudo opening	¡= 15-50 mi	in 15-50 mi	in 50-100 mi
Post	0.2	0.7	1.3	1.2	0.5	-0.5	-0.4	1.2	0.2	0.5	0.8
	(0.7)	(0.8)	(0.8)	(0.8)	(1.1)	(1.0)	(0.7)	(1.2)	(0.7)	(0.9)	(0.5)
Pat Lib x Post	3.2**	2.9*	2.4*	2.4*	1.9*	3.3*	2.4**	-0.5	3.2**	2.1	-1.4
	(1.5)	(1.5)	(1.2)	(1.2)	(1.1)	(1.7)	(1.0)	(2.4)	(1.5)	(1.9)	(1.4)
Mean Dep.	17.8	17.5	12.7	12.6	24.0	20.1	19.7	18.0	17.8	17.9	19.9
R2 (within)	0.12	0.11	0.10	0.10	0.16	0.19	0.45	0.15	0.12	0.26	0.33
Obs.	3432	3036	2486	2288	3894	2904	990	2926	3432	3432	3432

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:

$$\frac{\#Patents_{ijt}}{Population_{ij}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $PatLib_{ij}$ is an indicator if the library i is a patent library or if it is not, but belongs to the control observations of patent library i,and $Post_{it}$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use Federal Depository Libraries (FDLs) within the same state as controls. In column (2) we additionally use only FDLs that are also university libraries as controls if the patent library is also a university library. If the patent library is not a university library we only use FDLs that are not university libraries as controls. In column (3) we match on state, being a university and in addition we do a coarsened exact matching (CEM) with 5 bins on patent per capita in the year before the opening. In column (4) we add repeat the analysis of column (3) but also employ coarsened exact matching with 5 bins on population. Thus only FDLs in similar sized cities, with similar number of patent per capita and the same type of library are used as controls. In column (5) we do not match on state but take all FDLs within a 100 miles as controls. In column (6) we use the 5 closest FDLs as controls. In case FDLs are equally close to the treated library, we keep all of them. In column (7) we construct a counterfactual by keeping the share of patents around patent libraries among U.S. patents constant to pre-opening levels. In column (8) we assign a treatment indicator to the FDL closest to the patent library and drop all patent libraries from the sample. In column (9) we use the number of patents within 15 miles of the libraries as outcomes. This corresponds to our baseline specification. In columns (10) and (11) we use the number of patents between 15 and 50 miles and between t0 and 100 miles as outcomes, respectively. We use the weights suggested by Iacus, King and Porro (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.

Table B-3—: Further Robustness Tests

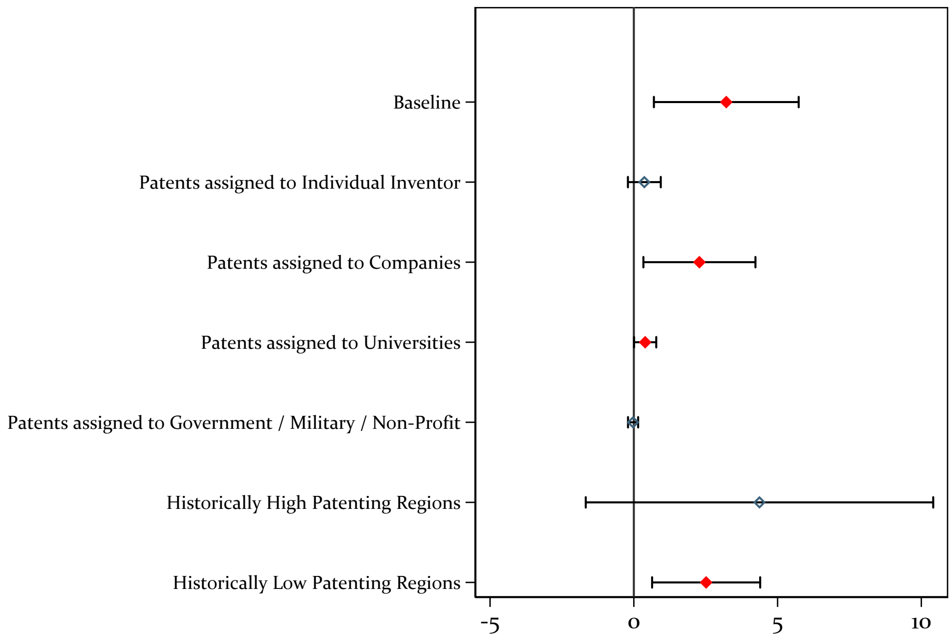
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline			Inventors who are..			# Patents	
	Baseline	W/o CEM Weights	PTDL spec. Trends	Co-located	25m around library	50m	# + Pop.	Ln(#+1) + Pop.
Post	0.2	0.0		0.3	-0.2	0.3	-1.4	0.0
	(0.7)	(0.7)		(0.6)	(0.5)	(0.5)	(3.6)	(0.0)
Pat Lib x Post	3.2**	3.3*	3.2**	2.5**	2.5*	2.7	31.3*	0.1***
	(1.5)	(1.7)	(1.5)	(1.1)	(1.3)	(1.7)	(17.7)	(0.0)
Mean Dep.	17.8	17.8	17.8	14.1	16.8	17.2	142.5	3.5
R2 (within)	0.12	0.12	0.40	0.08	0.23	0.33	0.22	0.38
Obs.	3432	3432	3432	3432	3432	3432	3432	3432

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:

$$\frac{\#Patents_{ijt}}{Population_{ij}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $PatLib_{ij}$ is an indicator if the library j is a patent library or if it is not, but belongs to the control observations of patent library i, and $Post_{it}$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In column (1) we use patents within 15m of the patent library per 100,000 inhabitants as the dependent variable. In column (2) we repeat our baseline estimation without using the weights suggested by Iacus, King and Porro (2012). In column (3) we include (assigned) patent library specific trends. In column (4) we only use patents by solo or co-located inventors within 15 miles from each other as the dependent variable. In the following two columns we use circles of 25 and 50 miles around the patent library to compute the dependent variable. In column (7), we use the number of patents in the 15 miles around the libraries as the dependent variable and control for (time-varying) population. In the final column, we repeat this estimation using the log number of patents as the dependent variable. In all regressions but column (2), we use the weights suggested by Iacus, King and Porro (2012) to identify the average treatment effect on the treated. Standard errors are clustered on the (assigned) patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure B-3. : Further Main Results



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

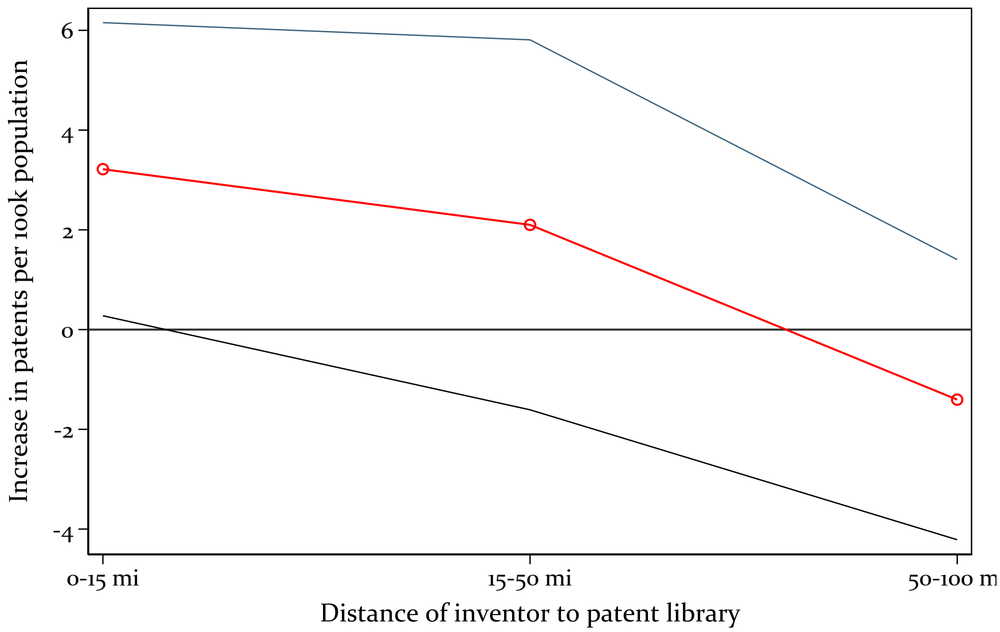
$$\frac{\#Patents_{it}}{Population} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_i \cdot Post_t + Library\,FE + Year\,FE + \varepsilon_i$$

where $PatLib_i$ is an indicator if the library i is a patent library and $Post_t$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. In the first line we report the point estimate for β_2 along with 90% confidence intervals. The confidence intervals are based on standard errors that are clustered on the patent library level. In lines (2) to (5) we split the dependent variable by the type of assignee. We show results separately for independent inventors, patents assigned to companies, patents assigned to universities, and patents assigned to the government, military or non-profits. In lines (6) and (7) we split the sample by an indicator if the region of the patent library has historically many or historically few patents. We define a region as having many patents if the average yearly number of patents per capita is above the median.

B5. Alternative Distances

In addition to not being present prior to patent deposit library opening, the effects that we find in (1) are not evident in regions outside of the patent library's commuting radius. Lines 9) to 11) of Figure 6 and Figure B-4 show that the increase in patents is largely localized in a geographic region most proximate to the arriving patent library. For patents filed by inventors whose addresses are further than 15 miles from opened patent libraries, the impact of library opening is of smaller economic magnitude and is not statistically significant. Beyond 50 miles, it is inexistent. In this analysis, we consider the outcome variable to be the number of patents in a variety of distance bands around the treatment and the control libraries. This result implies that the number of patents mostly increases around the patent library but not in the wider area. Further, the finding increases our confidence that regions are not receiving patent libraries in anticipation of increasing innovation potential. If a region was chosen to get a patent library based on an expected increase in its innovative capacity, the government must have been able pick exactly the right spot where patenting will increase.

Figure B-4. : 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:

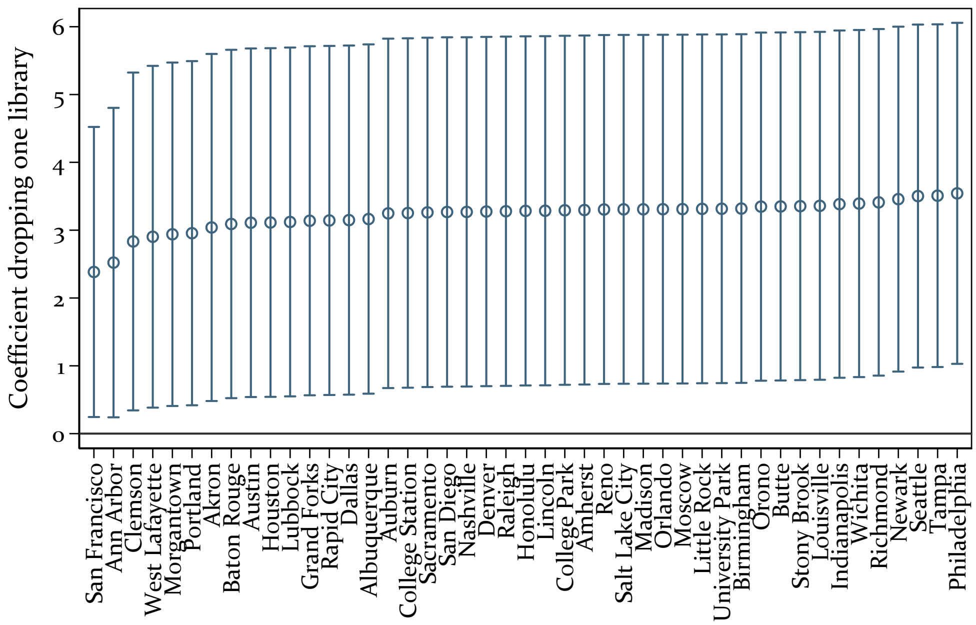
$$\frac{\#Patents_{ijt}}{Population_{ij}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $PatLib_{ij}$ is an indicator if the library j is a patent library or if it is not, but belongs to the control observations of patent library i, and $Post_{it}$ 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 only use patents in the distance band reported on the horizontal axis in the numerator of the dependent variable. We report 90% confidence intervals for the coefficient. The confidence intervals are based on standard errors that are clustered on the patent library level.

B6. Leave-one-out Estimation: The Impact of Individual Patent Depository Libraries

In our final set of analyses of the robustness of the results to alternative samples, we explore the role of individual library regions. In Figure B-5 we run our main analysis, dropping library regions one by one. With the exception of the library in Ann Arbor MI, we find that the coefficient indicating the post-patent library effect does not change. Dropping San Francisco reduces the coefficient from 3.2 to around 2.4, while making the estimate more precise though still within the initial confidence interval.

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



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:

$$\frac{\#Patents_{ijt}}{Population_{ij}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $PatLib_{ij}$ is an indicator if the library j is a patent library or if it is not, but belongs to the control observations of patent library i, and $Post_{it}$ 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.

As we described above, our main sample excludes the patent libraries of Burlington VT. This region has an extremely high patent per capita ratio because Burlington VT was the home of IBM's major research facility. This constitutes a substantial innovation outlier in its local area and, indeed, in the entire dataset. As a result, we could not identify a control region within 250 miles and within the same state that achieved even remotely similar levels of per capita patenting. When we add the library to our main analysis, we find a post library opening effect size greater than that in our preferred specification, but also that the additional noise renders the coefficient indistinguishable from zero.

B7. Structure of Patents: Citation Distance Increases, Patent Quality is Unchanged

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

To investigate this possibility, we compare bibliometric features of patents associated with inventors in patent library regions with control patents of the same technology field and the same filing year but that were filed by inventors in Federal Depository Library regions.

We again use the difference-in-differences specification in Equation (1), estimating now at the patent level, and asking how the nature of backward references and forward references change after library opening. We estimate each specification once for all patents of young companies (Panel A) and once for old companies (Panel B). We cluster standard errors at the patent library level.

Table B-4 reports the results of models assessing the impact of library opening on the nature of patents in affected regions for young and old companies. Column (1) shows that the average number of backward citations increased for young firms. Induced patents may, thus, have profited more from prior art. We explore the geographical range of patent citations in column (2) by examining how library opening affects the median geographic distance between citing and cited inventor. Patents of young companies experience an increase in backward citation distance. There is a similar but marginally insignificant effect for old companies in Panel B. These results are consistent with what we would expect if patent access for previously-inhibited inventors was the driving mechanism behind the core findings, i.e., that patent libraries improve the access to distant and therefore less likely to be known patents. The results in columns (3) and (4) suggest that patents produced after patent library opening are also more original, i.e., cite more technologically-distant prior art, although the effect is insignificant.[1]

One other issue worth exploring is the possibility that library opening does not induce innovation, but may simply cause a rush to submit any patentable invention. If library openings were to induce low quality patents, we would expect

[1]We define patent originality based on Hall, Jaffe and Trajtenberg (2001).

Table B-4—: Impact of Patent Libraries on Structure of Patents

	(1)	(2)	(3)	(4)	(5)
	Young companies				
	Backward citations	Median backward distance	Originality	# Fields	Forward citations
Pat Lib x Post	0.7	41.1*	2.2	9.9*	1.7
	(0.7)	(21.5)	(1.4)	(5.8)	(1.2)
Obs.	118649	118649	118649	118299	118649

	Old companies				
	Backward citations	Median backward distance	Originality	# Fields	Forward citations
Pat Lib x Post	-0.3	28.1	-1.0	-4.7	-0.6
	(0.3)	(20.3)	(0.7)	(4.4)	(0.7)
Obs.	175064	175064	175064	174441	175064

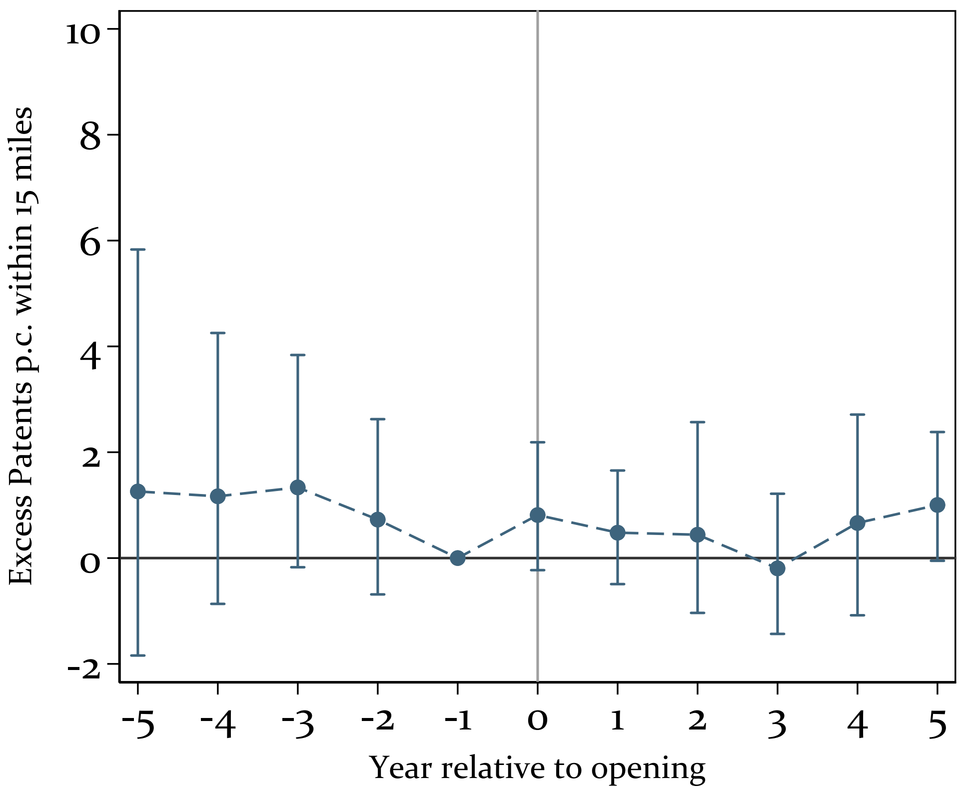
Note: This table shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period analogous to equation 1 but using other dependent variables: In column (1) we use the sum of backward citations. In column (2) we use the median distance between the location of the inventor of the cited patent and the citing patent j. In column (3) we use originality of the patent as defined by Hall, Jaffe and Trajtenberg (2001) and in column (4) we count the number of technical fields cited by the patent. In column (5) we use the sum of forward citations. The classification of technical fields follows Schmoch (2008). In column (1) we use a fixed effect for each combination of patent library, technology area and filing year as controls. In columns (2) to (5) we use a fixed effect for each combination of patent library, filing year and number of backward citations as controls. In Panel A we use only companies with their first patent less than three years before the opening of the patent library. An old company is a company with a patent more than three years before the opening. Standard errors are clustered on the patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

post-library patents to receive fewer forward citations than before. We investigate this question in column (5). The results evidence no decline in the number of forward citations, suggesting that induced patents are of similar quality (and value) to those produced before library opening. In sum, we interpret these results as consistent with the prospect that the mechanism behind the post-library patent boost is the improved access to previously-distant and expensive-to-access prior art.

B8. Patent Attorney Results

In Figure B-6 we use data from the historical rosters of registered patent attorneys at the USPTO (United States Patent and Trademark Office, 2018) to provide evidence on the impact of patent libraries on the local number of active patent attorneys. In line with our identification assumption, the number of patent attorneys in treatment and control group is similar prior the opening of libraries. In addition, there is no clear effect of the opening of patent libraries on the number of registered attorneys at the USPTO. These findings suggest that an influx of patent attorneys is unlikely to induce patent opening and is also unlikely to account for the boost in patenting that follows library arrival.

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



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

We also document the robustness of the paper's core results to controlling for the number of local patent attorneys in Table B-5. The first column replicates our baseline estimates. The second column does so for the subsample of library-year observations where patent attorney data is available. The third column shows that for this subsample, controlling for the number of patent attorneys per capita does not affect our estimates. If anything, this increases the estimated impact of opening a library. In line with what we would expect, the number of patent attorneys per capita in a region positively predicts local patenting. In the fourth column, we analyze the impact of opening a patent library on the number of patent attorneys per capita. In line with Figure B-6, patent libraries do not seem to affect the number of active patent attorneys.

Table B-5—: Patent Attorney Results

	(1)	(2)	(3)	(4)
	Dependent Variable			
	Patents p.c.			Attorneys p.c.
Post	0.2	0.3	0.4	-0.1
	(0.7)	(0.9)	(0.9)	(0.1)
Pat Lib x Post	3.2**	3.7**	3.8**	-0.1
	(1.5)	(1.7)	(1.6)	(0.2)
Patent Attorneys p.c.			1.0**	
			(0.4)	
Mean Dep.	17.8	18.3	18.3	1.6
R2 (within)	0.12	0.18	0.20	0.16
Obs.	3432	1949	1949	1949

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:

$$Outcome_{ijt} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $Outcome_{jt}$ is the number of patents per capita (columns 1-3) and the number of patent attorneys per capita (column 4) around library j that is filed in year t. $PatLib_{ij}$ is an indicator if the library j is a patent library or if it is not, but belongs to the control observations of patent library i, and $Post_{it}$ is an indicator for all years after the opening of the patent library. . As controls we use a fixed effect for each combination of patent library, technology class and filing year in columns (1), (2) and (4). In column (2), we only use those observations where patent attorney data is available. In column (3), we control for the number of patent attorneys per capita. Data on patent attorneys stems from the historical rosters of registered patent attorneys at the USPTO. Standard errors are clustered on the patent library level. *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Appendix to Section 5

C1. Example for Chemical Patents: Aspirin

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

Figure C-1. : Aspirin

UNITED STATES PATENT OFFICE.

FELIX HOFFMANN, OF ELBERFELD, GERMANY, ASSIGNOR TO THE FARBENFABRIKEN OF ELBERFELD COMPANY, OF NEW YORK.

ACETYL SALICYLIC ACID.

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

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

To all whom it may concern:

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

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

$$C_6H_4\left<\begin{matrix}OCO.CH_3\\COOH.\end{matrix}\right.$$

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

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

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

$$C_6H_4\left<\begin{matrix}OCOCH_3\\COOH\end{matrix}\right.$$

and exhibits therapeutical properties.

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

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

$$C_6H_4\left<\begin{matrix}O.COCH_3\\COOH\end{matrix}\right.$$

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

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

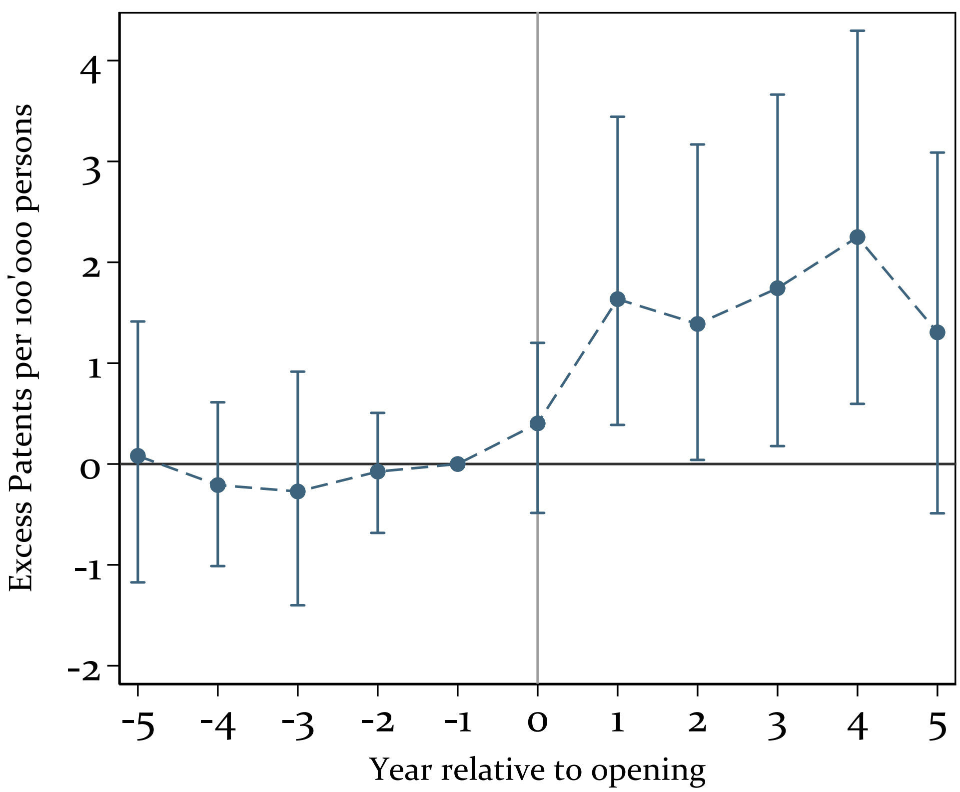
FELIX HOFFMANN.

Witnesses:
R. E. JAHN,
OTTO KÖNIG.

C2. *Time-Varying Treatment Effect for Chemistry*

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

Figure C-2. : Non-parametric Evidence: Chemistry

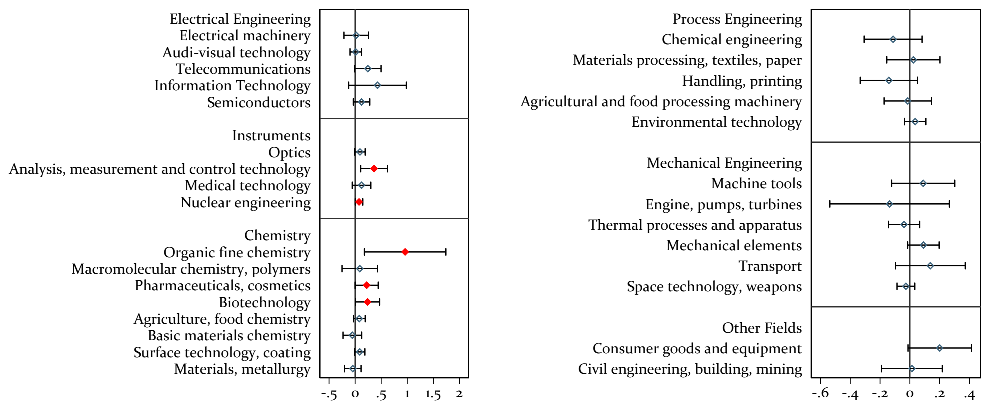


Note: This figure shows the yearly average treatment effects on the treated of opening up a patent library on the average number of patents in chemistry within 15 miles of patent libraries relative to the average number of patents around matched federal depository libraries. The 95% confidence intervals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

C3. Alternative Technology Classifications

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

Figure C-3. : Effect by Technology Category



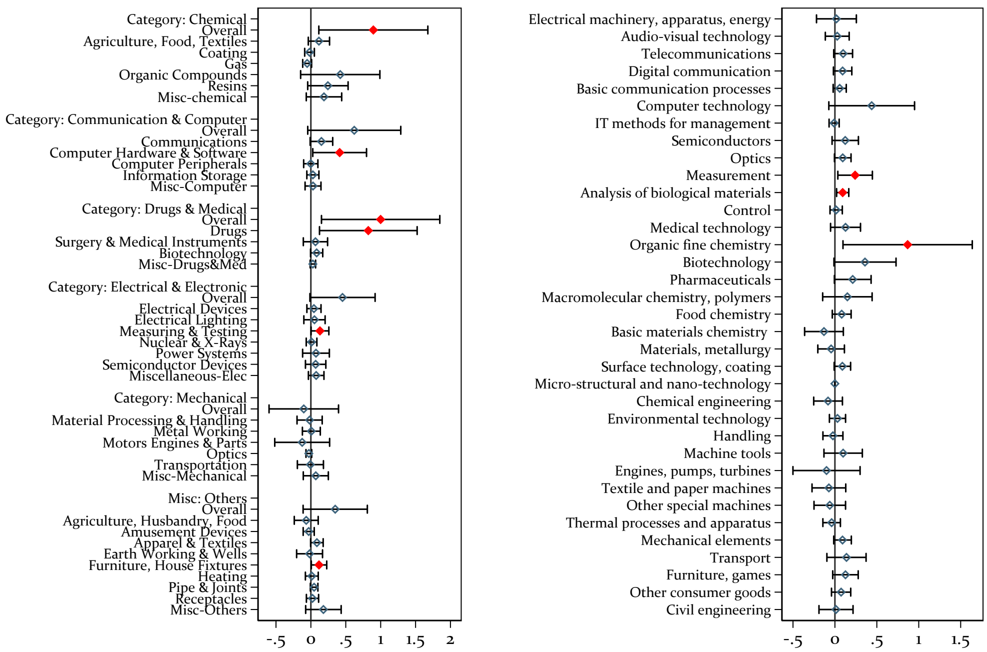
Note: This figure shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is:

$$\frac{\#Patents_{ijt}}{Population_{ij}} = \beta_1 \cdot Post_t + \beta_2 \cdot PatLib_{ij} \cdot Post_{it} + \alpha_{ij} + \gamma_t + \varepsilon_{ijt}$$

where $PatLib_{ij}$ is an indicator if the library j is a patent library or if it is not, but belongs to the control observations of patent library i, and $Post_{it}$ is an indicator for all years after the opening of the patent library. As controls we use library and year fixed effects. Each coefficient is from a separate regression where we split the dependent variable of our baseline regression by field as indicated in the figure. Thus, we only use patents in a specific field per 100,000 population as the dependent variable. The technological fields follow the ISI-OST-INPI classification of 1995 as defined in Schmoch (2008). The range plots indicate the 90% confidence intervals for the coefficient that are plotted with a hollow diamond if the coefficient is not significantly different from zero or a full diamond if the coefficient is significantly different from zero.

Figure C-4. : Alternative Technology Classifications

(a) By NBER Subcategory (USPTO) (b) ISI-OST-INPI Technological Categories 2008

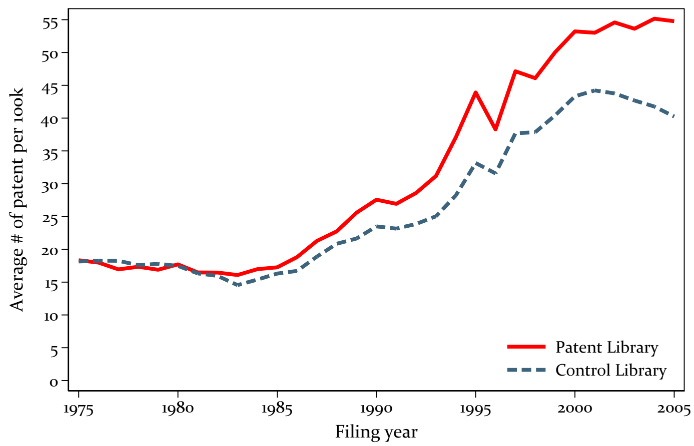


Note: These figures shows the results from a difference-in-differences estimation with five years before opening as pre-period and five years after opening as post-period. The estimation equation is analogous to equation 1. The technological fields in Subfigure a) are defined following the NBER Subcategories of Hall, Jaffe and Trajtenberg (2001) and in subfigure b) following the ISI-OST-INPI classification of Schmoch (2008). The range plots indicate the 90% confidence intervals for the coefficient that are plotted with a hollow diamond if the coefficient is not significantly different from zero or a full diamond if the coefficient is significantly different from zero.

C4. Long Run Effects of Opening a Patent Library

While patent libraries opened in the Internet era did not have the same impact on patenting as those opened in earlier periods, it is possible that the impact of earlier patent libraries was, nonetheless, long-lived. For example, it is possible that library opening and the concomitant boost in regional innovation may have improved the overall environment for R&D and commercialization, attracting new innovators and, potentially supporting a longer-term increase in innovative capacity (Audretsch and Feldman, 1996; Delgado, Porter and Stern, 2014). Figure C-5 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.[2]

Figure C-5. : Averages Over Time



Note: This figure plots the average number of patents per 100,000 population around (opened and not-yet opened) patent libraries (solid line) and their control libraries (dash line) over time. Thus, over time, an increasing number of libraries underlying the solid line are actually treated, while the control libraries never are. We delete the patent library in Burlington VT as in the main analysis.

[2]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.

*

REFERENCES

Arts, Sam, Bruno Cassiman, and Juan Carlos Gomez. 2018. "Text matching to measure patent similarity." *Strategic Management Journal*, 39(1): 62–84.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, 86(3): 630–640.

Balsmeier, Benjamin, Guan-Cheng Li, Tyler Chesebro, Guangzheng Zang, Gabe Fierro, Kevin Johnson, Aditya Kaulagi, Sonja Lueck, Doug O'Reagan, Bill Yeh, et al. 2018. "Machine learning and natural language processing on the patent corpus: Data, tools, and new measures." *Journal of Economics & Management Strategy*, 27(3): 535–553.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2014. "Clusters, convergence, and economic performance." *Research Policy*, 43(10): 1785–1799.

European Patent Office. 2016. "PATSTAT Global, Autumn Edition."

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. "The NBER patent citation data file: Lessons, insights and methodological tools." *NBER Working Paper No. 8498.*

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

Jenda, Claudine Arnold. 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.

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli. 2017. "Disambiguation of patent inventors and assignees using high-resolution geolocation data." *Scientific Data*, 4.

PatentsView. 2020. "PatentsView Database."

Schmoch, Ulrich. 2008. "Concept of a technology classification for country comparisons." *Final report to the world intellectual property organisation (wipo), WIPO.*

United States Patent and Trademark Office. 2018. "Historical Roster Books."