Knowledge Access: The Effects of Carnegie Libraries on Innovation*

Enrico Berkes[†] Peter Nencka[‡]

December 2021

Abstract

Between 1883 and 1919, Andrew Carnegie funded the construction of over 1,500 public libraries across the United States, reducing the costs of accessing knowledge for millions of people. We study the effect of these libraries on innovation using new data on city-level patenting and a novel control group: cities that qualified to receive a library grant and applied to be part of the program, but did not build a library. Patenting in recipient towns increased by 7-11 percent in the 20 years following library construction. We show that access to scientific knowledge and increased opportunities to collaborate are possible mechanisms.

Keywords: Access to Knowledge, Innovation, Libraries.

JEL Classification: H4, L3, N9, O3, R1.

*Thanks to Meta Brown, Jeffrey Furman, Ruben Gaetani, Joshua Hall, Walker Hanlon, Lauren Jones, Ezra Karger, Kurt Lavetti, Dean Lillard, Trevon Logan, Martin Saavedra, Marianne Wanamaker, and Bruce Weinberg for their helpful comments. We also thank seminar participants at Ohio State; West Virginia University; University of Pittsburgh; NBER Summer Institute; Università della Svizzera Italiana; Max Planck Institute for Innovation and Competition; European Meeting of the Urban Economics Association; Cliometrics Society; Society for Economic Dynamics; North American Meeting of the Econometrics Society; Western Economic Association; the North American Meeting of the Urban Economics Association; University of Toronto; Dickinson College; University of Richmond; Miami University; EPFL Virtual Innovation Seminar; NBER Productivity Seminar; New York University; and the Economic History Association for their comments. Tiger Sun provided excellent research assistance. We thank Luca Bittarello for sharing his Knights of Labor and labor strike data.

[†]The Ohio State University, berkes.8@osu.edu.

[‡]Miami University, nenckap@miamioh.edu.

And daily in the papers thou shalt read,
Of ten new libraries, in cities vast,
In villages, and Indian wigwams too,
In Texas ranches and Esquimaux huts,
In Heaven, Hell, and stations in between
– Upton Sinclair, 1902 poem

1 Introduction

The recombination of existing ideas is a key component of the innovation process (e.g., Weitzman, 1998). Isaac Newton famously declared that his work was built "by standing upon the shoulders" of past scientific thinkers. For this reason, researchers and policymakers alike believe that expanding access to knowledge may promote innovation. Mokyr (2002) argues that the spread of institutions that reduce the costs of accessing knowledge—such as national science academies and scientific journals—contributed to the outset of the Industrial Revolution in Britain. Even today, governments hope that expanding high speed internet and cell networks in developing countries might increase innovative output by broadening access to knowledge (e.g., United Nations, 2018).

However, causal evidence that broadening access to knowledge increases innovation is scarce. Researchers who want to study this link typically face two major hurdles. First, many institutions that disseminate knowledge only do so in narrow ways or are targeted towards specific groups, such as scientists (e.g., Bryan and Ozcan, 2021). Second, institutions that spread knowledge (e.g., colleges) can have other simultaneous effects, like attracting high-skilled immigrants, which makes it difficult to isolate the effects of access to knowledge from other factors that might also affect innovation (e.g., Andrews, 2021b).

In this paper, we study the rapid rollout of an institution that dramatically lowered knowledge access costs for millions of people: the local public library. We estimate the effects on innovation of more than 1,500 new, high-quality U.S. public libraries financed by the steel titan and philanthropist Andrew Carnegie between 1883 and 1919. For residents of recipient towns, a Carnegie library was a new—and often the only—local source of scientific, technical, and practical knowledge. The library provided citizens with new opportunities to exchange and access new ideas thanks to books and magazines. Libraries also provided opportunities to interact with patrons with similar interests.

The characteristics of public libraries make them particularly suited to studying the relationship between access to knowledge and innovation. First, public libraries are open to all, regardless of income or social class. This contrasts with other information-spreading institutions that are only accessible to subsets of the population. The public nature of libraries allows us to study the effects of increased information on potential "lost Einsteins" (Bell et al., 2018)—groups underrepresented in innovative and scientific outputs, such as women, different ethnic groups, and those who live in rural areas. Second, libraries are local institutions. Evidence shows that knowledge spillovers are relatively local and sharply decay over space (e.g., Jaffe et al., 1993; Murata et al., 2014). The historical rollout of libraries provides variation in information access across otherwise similar, narrowly-defined geographic areas. Finally, libraries are unlikely to change the local innovation environment in ways that are unrelated to information access—such as attracting new businesses—allowing us to isolate the effect of knowledge access.

To identify the causal effect of Carnegie libraries on innovation, we estimate difference-in-differences models that exploit the sharp timing of library construction. We leverage a wealth of institutional information that we gather on the Carnegie program to construct a novel control group. We identify more than 200 cities that applied for the program, qualified to receive a library grant, received preliminary construction approval, but ultimately did not execute the project. Many of these cities rejected Carnegie's donation due to his unpopularity, particularly among labor associations. We show that these cities are similar to cities that eventually built a library along various demographic and economic characteristics. Importantly, the two groups of cities also follow parallel patenting trends prior to library entry.

We find that patenting increases in towns that built Carnegie libraries relative to control cities. Patenting starts diverging shortly after receiving a library grant, a pattern which is consistent with the typical construction times observed in the data. Differences between cities that accepted and rejected libraries peak between 5 and 15 years after acceptance. Overall, patenting in cities that accepted Carnegie libraries increased by approximately 7-11 percent in the 20 years after library entry. We show that our findings are not driven by an increase in low-quality patenting or an increase in city population. Both women and foreign-born patented more after libraries opened, suggesting that public libraries helped expand access to knowledge for groups that were underrepresented in patenting, although their relative contribution to patenting remained largely unchanged. We demonstrate that our results are robust to a range of sample, measurement, and estimation choices. To the best of our knowledge, we are the first to provide estimates of the effect of public libraries on innovation.

We propose the existence of two mechanisms that could explain the link between libraries and patenting that we describe. First, we test whether access to new library materials contributed to patenting increases. To do this, we separately estimate effects by patent technology classes. We find that libraries had the largest impact on classes related to practical trades, such as farming, construction, and mechanical engineering, consistent with library collections. We observe smaller effects in highly technical fields, such as chemistry and physics, where innovation likely required more human or physical capital. To further explore this mechanism, we test whether patents in treated cities were more likely to cite books, prior patents, or magazines. We compile a list of keywords commonly used to cite these materials and identify patents that contain these keywords in their text. We find suggestive evidence that libraries increased both the probability of observing a patent that cites prior work and the number of such patents.

Second, we test whether creative collaborations increased after libraries opened. We provide evidence for this social channel by estimating the effect of Carnegie libraries on patents authored by multiple inventors. Like today, libraries were a central gathering point for meetings and events in the early 20th century. For example, in the 1920s, the Carnegie library in Cleveland, Ohio had a robust series of social reading groups. In St. Louis, Missouri, a variety of groups met at the library, such as "school clubs, groups of foreigners of many nationalities, women's clubs of all kinds, mothers' clubs, parliamentary classes, socialists, religious meetings, dance clubs and classes, political clubs and meetings, and musical organizations" (Learned, 1924). We find that multi-inventor patents increased after libraries opened.

Related Literature. This paper relates to three strands of the literature. First, we provide new evidence on how *broadening* access to knowledge affects innovation. Prior work has largely focused on the effects of expanding information access through the patent system or scientific literature, likely impacting innovators and scientists. For example, a set of papers finds that patent disclosures facilitates future innovation (e.g., Graham and Hegde, 2015; Hegde and Luo, 2018; Gross, 2019) and Furman et al. (2021) finds that patent deposit libraries increased local patenting. Iaria et al. (2018) find that shocks to international scientific cooperation during World War I reduced productivity for scientists who relied on foreign research. Biasi and Moser (2021) find that stripping copyrights from German scientific books in 1943 led to price declines for scientific books and a subsequent increase in the probability of citing those books in scientific articles, patents, and PhD theses. We contribute to this literature by exploiting a distinct shock to information access caused by an easily accessible institution that is open to all. Our results provide a historical complement to recent work on the information-spreading power of the internet and websites like Wikipedia (e.g., Czernich et al., 2011; Cardona et al., 2013; Akerman et al., 2015; Thompson and Hanley, 2018; Derksen et al., 2019; Xu et al., 2019).

Second, we contribute to a literature studying how local institutions affect innovative activity. Previous studies in this literature have mainly focused on colleges (e.g., Furman and MacGarvie, 2007; Aghion et al., 2009; Kantor and Whalley, 2014; Hausman, 2020; Kantor and Whalley, 2019; Andrews, 2021b). These papers consistently find that after the establishment of a college, innovative activity increases, although they disagree on the channels through which this effect operates, as well as its magnitude. Relative to colleges, public libraries played a distinct role in disseminating information in the early 20th century. Libraries provided low-cost, state-of-the-art information across disciplines, as well as information on the patenting process itself.

Third, we contribute to an emerging literature on the impact of public libraries in the United States. In addition to the work of library historians who have chronicled Carnegie's programs (e.g., Bobinski, 1969; Jones, 1997), our work relates to recent papers on the political economy and development of libraries. Kevane and Sundstrom (2014) outline the characteristics that predicted local library entry in the early 20th century, including the positive impact of state library associations (Kevane and Sundstrom, 2016b). Kevane and Sundstrom (2016a) estimate the effect of library entry during the 20th century on short-run political outcomes. They find no clear relationship between library entry and participation in the following election. We expand this work by studying human capital and innovation outcomes and introducing a new control group. In work that began contemporaneously with this project, Karger (2021) estimates the effects of Carnegie libraries on long-run schooling and occupation choices and finds that libraries positively affected education and occupational upskilling. We view our projects as complements, given our focus on the right tail of the distribution of human capital (i.e., inventors). In a more modern context, Gilpin et al. (2021) study the effects of library capital spending during the 2000s on library use and child test scores. They find that library investment increased library quality, library usage, and the reading test scores of nearby children.

The rest of the paper is organized as follows. Section 2 overviews the history of Carnegie's library program and reviews historical examples of inventors using public libraries. Section 3 discusses the construction of the data. Section 4 presents the empirical strategy and discusses the main results. Section 5 proposes and explores possible mechanisms and discusses heterogeneity across inventor and city characteristics. Section 6 shows that our results are robust to a variety of specification, measurement, sample, and estimation choices. Section 7 concludes.

2 Historical background

In this section, we describe the history of public libraries in the United States during the 19th century as well as the details of the Carnegie library program. We then discuss the nature of innovation at the time of Carnegie's grants and provide qualitative evidence showing links between innovation and libraries during this time in the United States.

2.1 Public libraries before Carnegie

Public libraries in the United States are a relatively recent civic institution. In 1833, the small town of Petersborough, N.H., established the first U.S. library open to all citizens and supported by town tax dollars. In the early 1850s, New Hampshire, Maine, and Massachusetts passed state laws authorizing local taxation to finance libraries. These laws helped spread public libraries in Northeastern cities. The first large city to open a municipal library was Boston in 1852. Despite a growing movement in favor of public libraries, their diffusion during the rest of the 19th century was slow. The financial pressures of the Civil War reduced resources for publicly funded libraries. It was not until 1893 that the growth of libraries hit a turning point thanks to the Chicago World Fair. There, the American Library Association (ALA)—an interest group of librarians founded in 1876 that advocates for the spread of municipally-funded libraries—showcased a demonstration public library with 5,000 books. Spearheaded by ALA's president, Melvil Dewey, the exhibit attracted national attention (Sharp, 1893; Wiegand, 2015). The popularity of this exhibit helped fuel the local demand for public libraries.

In the early 1890s, at the time of the ALA exhibit and at the cusp of the library revolution, there were approximately 600 public libraries nationwide, almost exclusively located in the Northeast (Jones, 1997). Outside large cities like Boston, many of these libraries were not freestanding, but instead located in the basements or attics of pre-existing buildings. For example, in Malta, Mont., the library was located on the balcony of a drugstore; in Dunkirk, N.Y., in the basement of a hospital; in Marysville, Ohio, in the horse stall of the fire department (Bobinski, 1968). Less than 30 years later in 1919, the United States had 3,500 public libraries, most in dedicated buildings and many occupying the largest building in town. Over half of these new libraries were constructed with funds from a single donor: Andrew Carnegie.

2.2 The Carnegie library program

Andrew Carnegie's library funding program is one of the most wide-reaching acts of philanthropy in U.S. history. From his first grant in 1883 (to Allegheny, N.Y.) to his last grant in 1919 (a branch library in Philadelphia, Pa.), Carnegie fully funded the construction of 1,687 public library buildings across the country at the cost of approximately $1 billion in 2020 dollars. Carnegie's stated motivation for the library grants was consistent with his larger views on philanthropy: He believed that public libraries were a way citizens could improve themselves if they had sufficient drive.

Carnegie's library grants started with a small number of cities, but quickly became a national phenomenon. Carnegie himself referred to two distinct periods of his program: the "retail" and "wholesale" phases. In the "retail" phase that started in 1883, Carnegie gave money to build eight libraries in selected communities.[1] By 1899, Carnegie shifted his priority to providing library access for as many people as possible (the "wholesale" phase). He opened the library application process to essentially all cities that did not already possess a stand-alone, self-sufficient library.[2] In some cases, cities already had small libraries that the Carnegie program supplanted. In this case, our estimated treatment effects of Carnegie libraries will not solely reflect the effects of *new* libraries; instead, it will pick up the effects of new, *high-quality* libraries.[3]

[1]These locations often has personal ties to Carnegie. For example, Carnegie's earliest U.S. library was built in Allegheny, PA, near one of his steel mills.

[2]In our baseline empirical analysis, we exclude "retail" libraries, since their hand-picked nature may imply that selected cities were systematically different from other cities. In the robustness checks, we show that our results are robust to including these libraries in our sample. Carnegie also required cities to have over 1,000 people. However, evidence shows that this requirement was not binding throughout the period considered here. In fact, we observe several cities which got a grant although their population was below 1,000. Most cities were much larger than 1,000, but we also observe cities smaller than 1,000 that received a library (e.g., Bayliss, California) in conjunction with support from counties or nearby townships.

[3]Unfortunately, temporally consistent library data is difficult to compile for the universe of cities in the United States. While this information was collected by the U.S. Census, the censuses were not regular and had shifting inclusion requirements across years. For that reason, we focus on identifying the net effect of Carnegie libraries, which involves

The grant application process started with a letter of interest to Carnegie's private secretary and administrator of the library program, James Bertram. Often initial letters to Bertram came from everyday citizens or leaders of civic groups. Bertram would reply, noting that he was happy to hear about their interest in libraries but that future correspondence should occur with elected city representatives. Bertram instructed cities to fill out a short form, which asked for information on the city population, the names of city officials, whether the city already had a public library, and if so, for additional details on its building structure, its expenses, and its circulation. This step ensured that cities understood what Carnegie was willing to supply—money to construct the building for a new public library—and identify whether such a library already existed. Almost all cities that applied were found eligible and progressed to the next stage.[4]

This next stage required that cities satisfy a number of requirements in order to receive construction funds. Accepted libraries received a short letter from Bertram, like this one to Stoneham, Mass. reported in Jones (1997):

> Dear Sir:
>
> Responding to your communication on behalf of Stoneham. If the City agrees...to maintain a Free Public Library at a cost of not less than Fifteen Hundred Dollars a year, and provides a suitable site for the building, Mr. Carnegie will be glad to furnish Fifteen Thousand dollars to erect a Free Public Library Building for Stoneham.

The letter highlights the award amount that Carnegie judged necessary to fully construct the library. It also outlines four main features of Carnegie's grants:[5]

1. **The granted amount was determined by Carnegie and Bertram.** With few exceptions, Carnegie and Bertram decided the exact grant amount based on reported population. The rate was approximately $2-3 per person in the town. The grant amount could be controversial. Some cities argued that they were entitled to additional funds, often by noting that the census

both quality and quantity dimensions.

[4] As noted in Bobinski (1969), Carnegie and his staff rejected requests at this stage for state, subscription, and historical society libraries. Carnegie also funded the construction of 108 academic libraries during this period. We do not consider these libraries in this paper, as they had a distinct mission from public libraries and may have not been open to the general public.

[5] For more details on the structure of the program, see the excellent histories in Bobinski (1969) and Jones (1997). After 1908, Carnegie began to impose more requirements on the specific construction techniques and floor plans that libraries could use and cities were required to submit blueprints for approval. This occurred after a number of towns tried to combine libraries with other civic buildings, like gyms or city halls, that Carnegie was not interested in funding.

population figure was out of date and that they expected to draw attendees from beyond their city limits. These protests rarely succeeded (Bobinski, 1969).

2. **Carnegie libraries needed to be free and public.** Carnegie libraries were meant to be open to the public, unlike many private libraries of the day, and were not supposed to charge admission fees, unlike commercial, subscription libraries.

3. **The construction site needed to be provided by the city.** Carnegie required that the city either purchase a site or re-purpose an existing city property. Bertram asked cities to send proof of site ownership before the funds were dispensed. Bobinski (1969) estimates that one in three cities had some sort of controversy about the site location. Because libraries often became town centerpieces, it is unsurprising that citizens fiercely argued in favor of their preferred locations.

4. **Cities were required to commit funds for ongoing maintenance of the libraries.** Carnegie knew that providing funds for construction was not enough. He wanted to make sure that cities could fill the libraries with books, pay the staff, and maintain the building. His solution, as illustrated in the above letter, was to require that cities pledge to spend 10 percent of the initial construction grant on annual library upkeep. Practically, this 10 percent maintenance requirement was at the lower end of what would be required to staff and maintain a library in the early 20th century. Cities often had to allocate additional funds beyond the 10 percent to keep their libraries running, particularly as average city sizes grew throughout the 1920s and 1930s (Bobinski, 1969).

   Despite the written pledge, once a library was built, Carnegie had little ability to enforce the 10 percent contribution requirement. Cities knew that enforcement was limited, and there is ample evidence of cities failing to meet the requirement. In 1917, the Carnegie Corporation—which by this point had been founded to manage the library program and related philanthropy—sent a survey to investigate reports that the pledge was not being met. The results were stark and discouraging. In Ohio, for example, 23 out of 77 cities were not meeting the pledge (Bobinski, 1969). After this non-compliance was discovered, library grant-giving to Ohio was briefly suspended, but there was no direct action against the offending libraries themselves.

2.3 Reactions to Carnegie's library program

The reactions to Carnegie's program were mixed. Many cities welcomed Carnegie's money. Indeed, communities that received a library could create a cascading effect within a state, as residents in nearby cities rushed to apply for their own building. But for some residents, Carnegie grants were controversial. A large number of cities that would have qualified to build a library never applied. In addition, more than 200 cities applied for and were granted funds but ultimately rejected the grant. The decision of these cities, reflecting approximately 15 percent of offered grants, was a notable rejection of Carnegie's program. Throughout the rest of the paper, we refer to these cities as *rejecting cities*.

The key event that generated Carnegie's long-term negative reputation and many eventual rejections was the steel worker strike at Homestead, Pennsylvania, in July 1892. After months of rising tensions in the face of increasing production demands by Carnegie's managers, the local union and management were unable to reach an agreement on a new contract. Determined to defeat the union, Carnegie's factory locked out the union workers, and workers struck. Carnegie's managers hired a private militia to break the strike and take back the town.[6] The resulting battle led to the deaths of nine strikers, ten members of the militia, and scores of wounded. The battle made international news. Carnegie's actions were never forgotten by those in the labor movement, many of whom later became involved in the fight against libraries.

An editorial published 17 years after the strike in 1909 in the *Pittsburg Kansan* illustrates the long-lasting impact of the strike and Carnegie's unpopularity among labor in library debates:

> A library that is built on money wrung from the hearts and homes of Homestead miners who were shot down in cold blood...is no fitting monument for the kind of men that built Pittsburg. If Mr. Carnegie wants to be charitable, let him commence with the widows and orphans of the murdered miners. (reproduced in Jones, 1997)

In Wheeling, West Virginia, which ultimately rejected Carnegie's offer of funding, a union leader declared that "[i]n view of Mr. Carnegie's attitude toward labor it is the duty of organized labor to adopt stringent measures to defeat the erection of this disgraceful monument" (Electrical Worker,

[6]Whether Carnegie knew of the exact actions taken by his managers and how much blame Carnegie should face for the resulting deaths has been long debated by historians and biographers. For a recent discussion of the event, see Nasaw (2007).

1901). In Detroit, where opposition to Carnegie was fierce, the city treasurer proclaimed "[w]e ought to be able to take care of ourselves…[not] accept a big chunk of money as a gift from a man who has made his money the way Carnegie did" (Krass, 2011). Opposition was not limited to local officials. Prominent national politicians and writers, including socialist and recurring presidential candidate Eugene V. Debs and *The Jungle* author Upton Sinclair, spoke out against accepting libraries. Even Samuel Clemens (better known as Mark Twain) weighed in, noting that Carnegie's quest for personal recognition might be behind his generosity: "He bought fame and paid cash for it" (Bobinski, 1969). While labor sentiment could drive Carnegie rejection, it was not always dispositive and likely idiosyncratic. For example, Homestead, PA itself built a Carnegie library, and the president of the American Federation of Labor, Samuel Gompers, famously stated: "After all is said and done, he might put his money to a much worse act. Yes, accept his library, organize the workers, secure better conditions and particularly, reduction in hours of labor, and then workers will have some chance and leisure in which to read books."

Opposition from the political left was only one of the obstacles standing between cities and their libraries. The 10 percent yearly pledge was also unpopular, despite being at the lower end of what was needed to support a library as well as being largely unenforceable. In order to avoid the 10 percent pledge, some cities rejected Carnegie and either did not build a library or courted local philanthropists instead.[7] In addition, some cities could not secure or decide on a library site, eventually forfeiting their application.

2.4 Innovation during the Second Industrial Revolution

Having described the Carnegie library program itself, we now turn to evidence that suggests a potential relationship between libraries and innovation. Both the characteristics of innovation during the early 20th century and the contemporaneous records of inventors themselves suggest that libraries could have played an important role in the idea-generating process at this time in U.S history.

Patent data show that inventions claimed between 1870 and 1930 were often simple, concentrated in practical technology classes, and usually claimed by a single inventor. Figure 1 shows the share of single-authored patents filed from 1870 onward. Until 1930, about 90 percent of all patents issued by the U.S. Patent and Trademark Office (USPTO) fell in this group. Since then, this share has

[7] We exclude cities that built libraries from local philanthropists from our baseline sample.

steadily declined, reaching about 35 percent in 2010. Figure 2 plots the distribution of patents across technology classes by decade. This chart shows that during the Carnegie library expansion period, patenting activity was mainly concentrated in relatively practical trades, such as human necessities (which includes farming), performing operations and transporting (e.g., vehicles, metal casting), and mechanical engineering. About 70 percent of the patents filed between 1870 and 1930 belonged to these three technology classes. Electricity, physics (which includes computing), and chemistry patenting only expanded in more recent decades.

In this context, it is plausible that public libraries were associated with increases in innovative activities. Capital-intensive, technologically demanding, team-based innovation that we associate with modern patenting was less prevalent. Instead, much innovation was done individually or in small teams and focused on trades that overlap with the types of practical materials that libraries held.

2.5 Documented relationships between public libraries and innovation

Biographies of inventors also provide direct evidence that libraries affected creative and innovative output. These anecdotes do not establish the causal effect of libraries, but they do provide context for the results that we identify and suggest potential mechanisms.

In the 1930s, Chester F. Carlson—the inventor of the modern copy machine and founder of Xerox—cited research in the public library as a key input in his work:

> At first, I did as much thinking as I could about the problem. I jutted down my thoughts in my inventor's notebook. But mainly, in the beginning, I started reading. I know I spent many evenings and weekends in the Science and Technology Division of the New York Public Library. I got out everything I could find on printing and duplicating. (Researching NYC, 2015)

Edward Land, who invented the Polaroid method of instant photography, also made breakthroughs in his research at the New York Public Library in the 1920s. The key sources that led to both Carlson's and Land's breakthroughs in the library were over 30 years old when they found them (Wenyon, 2009). Because libraries recorded and organized both current and past technical materials, they provided a unique opportunity to discover and recombine existing ideas.

The iconic design of the Coca-Cola bottle was also developed with help from a public library. In 1915, Coca-Cola sent out a call for bottle designs to differentiate their product from competitors and counterfeiters. Designers at the Root Glass Company in Terre Haute, Indiana visited their local public library for research. They were inspired by an illustration of a distinctive cocoa bean and patented their design. Their bottle was adopted by Coca-Cola in 1916 (The Coca-Cola Company, 2015).

Even Thomas Edison made regular use of public libraries to access knowledge on electricity and telegraphs that he subsequently used in his inventions. As written in a recent biography:

> In the library stacks, Tom tracked down Dionysius Lardner's classic work on the *Electric Telegraph* as well as his *Handbook of Electricity, Magnetism and Acoustics*. He read Richard Culley's *Handbook of Practical Telegraphy*, Charles Walker's *Electric Telegraph Manipulation*, and Robert Sabine's *History and Practice of the Electric Telegraph*. (Baldwin, 2001)

Finally, a 2012 obituary for Stanford Ovshinsky, a self-taught inventor whose work on batteries transformed the industry, highlighted the long-term importance of his hometown Carnegie library in his development:

> A mediocre student, he spent hours in the Akron public library, where his real education took place. 'His teachers didn't understand him, but his librarian did,' his son said Friday. (The Los Angeles Times, 2012)

These examples suggest that public libraries played a role in the development of iconic innovations in the early 20th century.

3 Data

In this section, we describe the data that we use in our empirical analysis. We first describe our historical data on libraries. We then explain the construction of our patenting data, which is the source for most of our outcome variables. Last, we discuss other city-level covariates and provide summary statistics. Further details on data construction are provided in Appendix B.

3.1 Accepted and non-built libraries

We construct a dataset of all Carnegie libraries using historical records collected by Bobinski (1969) and Jones (1997). Both authors compiled their lists from the original Carnegie library program correspondence and surveys of libraries.[8] We assign each library to the city where it was built and use grant years as "treatment" years since they exist for cities that both accepted and rejected grants. In some larger cities, Carnegie funded multiple libraries. These awards typically paid for the construction of the main library building, as well as branch libraries. Multi-library grants occurred in approximately 5 percent of recipient towns and accounted for roughly 200 of the 1,687 Carnegie grants. In multi-library cities that built libraries at different points in time, we assign the city-level grant year to the first time that a grant was offered.[9]

We also collect information on when each library opened. We compile opening dates from library websites, phone calls to libraries, and architectural historical society records.[10] With these data, we calculate the average time to library opening after grant dates and confirm that the timing of our results is consistent with construction patterns. Figure A1 shows the distribution of construction times. On average, libraries took three years to open after grants were made. But the majority were built and opened to the public within 1-2 calendar years after the grant being disbursed.

To identify cities that rejected Carnegie library grants once they were approved, we rely on Bobinski (1969). Bobinski identified 209 Carnegie libraries "that never materialized." His primary source is the original Carnegie library correspondence between Bertram and rejecting cities, which we have also requested and reviewed. In addition to the locations of the rejecting cities, Bobinski identifies the grant amount and the date of the offer.

From the universe of library grants, we construct a consistent sample of cities that is used throughout the main analysis. First, we exclude a handful of library grants that Carnegie made before 1899 during the "retail period." Carnegie hand-selected the location of these grants before he opened his program to national applications and had personal connections to these early locations, so including them may introduce bias into our analysis. Second, we exclude Carnegie grants in larger cities and counties. Carnegie grants in larger cities were distinct and came with additional

[8] We also completed a careful review of the original Carnegie correspondence and news archives to search for any records that these authors might have missed—we did not identify any additional libraries.

[9] As discussed below, we omit most of these large cities from our baseline analysis. Our results are robust to alternative date choices, including assigning the median granted year.

[10] This data was compiled and verified in conjunction with Ezra Karger, who uses the data in Karger (2021).

requirements and benefits relative to the vast majority of other cities. Moreover, many of these cities and counties (e.g., Cook County, Illinois) already had well-developed library systems that Carnegie improved.[11] Third, we exclude control cities that rejected Carnegie because they instead built a library with funds from a local philanthropist. Finally, we exclude cities that cannot be matched to the 1900 census, which we use as our main data source of pre-program city characteristics. These cities were often unincorporated in 1900 and their city names were not recorded by census enumerators; this was particularly common in the Western United States. We show in Section 6 that these restrictions do not affect our main conclusions.[12]

3.2 Patent data

We measure innovative activities at the local level through patent data. Patents are a popular but imperfect measure of innovation: not all inventions are patented, and the propensity to patent is a function of the underlying legal environment (e.g. Moser, 2005). In our setting, patents are the only consistently available measure of local innovation at the local level, which is key given the local nature of our treatment. Further, patent data allows us to investigate the underlying content of inventions.

Our patent data come from the Comprehensive Universe of U.S. Patents (CUSP; Berkes, 2018). The CUSP covers more than 90 percent of the patents issued by the U.S. Patent and Trademark Office (USPTO) over the period 1836-2015.[13] From this dataset, we collect information on the distribution of technology classes associated to each patent according to the Cooperative Patent Classification (CPC), inventors' names, filing year, the raw text of patents, and inventors' cities of residence.

With these data, we create a longitudinal dataset of patents by filing year in each city that accepted or rejected a Carnegie library. When there are multiple authors, we proportionally assign patents to cities by assigning to each city a fraction of the patent equal to the inverse of the number of authors.[14] Each patent is associated to a distribution of technology classes according to the

[11] Our baseline sample is constructed by excluding all cities that had more than 30,000 people and all counties with more than 750,000 people. We discuss how our results change when using alternative cutoffs and samples in subsequent sections.

[12] A full description of how each sample restriction affects our sample size and estimated patenting impacts is shown in Table 10

[13] Berkes (2018) provides a full description of these data and how it was compiled. Andrews (2021a) compares the CUSP with other existing patent data sources, and suggests that the CUSP is currently the "gold standard" for researchers interested in patent and inventor-level information.

[14] As shown in Figure 1, approximately 90 percent of patents were single-authored during the time frame of our

Cooperative Patent Classification (CPC).[15]

We assign two measures of quality to each patent. First, we calculate the number of forward citations. Since patents filed before 1947 lack a reference section, patents in our sample were not required to cite prior art. However, patents from the 1870-1930 period are often cited by later patents once the citations section was introduced. As is common in the literature, we use citations as a positive indicator of invention quality. We use both the count of citations and the probability of observing a patent with at least one citation in a given city-year as measures of quality. Second, we use the measure of patent quality recently developed by Kelly et al. (2021). Kelly et al. (2021) use the text of patents to measure their impact by comparing the text of a given patent to the text of those that came before and after it. According to this measure, a high quality patent is one that is dissimilar to prior patents but similar to future ones. That is, it is a patent that is both novel for its time and shapes the direction of future innovation. We identify patents in the top ten percent of Kelly et al. (2021)'s baseline quality measure and test whether libraries increase the probability of observing one of these patents.[16]

For every patent, we also predict the likelihood that each author listed is a female or an immigrant. From the patent data, we observe the full name of inventors. To assign gender and immigration probabilities, we use the 1900-1940 full count census files. For each unique first name in the census, we calculate the proportion of respondents who are female, and we assign this proportion to each corresponding inventor name.[17] We perform a similar exercise for immigrants, using the last name and country of origin of people in the census.[18]

To test for potential mechanisms, we identify the subset of patents that cite books, prior patents, or magazines. We do so by specifying an initial list of keywords that are likely to be associated to those materials (e.g., encyclopedia, handbook, dictionary, etc.). We search the corpus of patents and identify those that mention the keywords. We then manually review our matches to identify other

analysis. Many of the remaining 10 percent are co-located authors. Assigning full credit for a patent to every authors' city (instead of dividing by $1/n$) does not change our results.

[15]Similarly to what we do for multi-authored patents, each patent which is assigned to multiple classes is proportionally distributed across classes.

[16]Kelly et al. (2021) discuss and develop a number of closely related patent quality measures. Our results using the top 10 percentile method are similar if we instead use their alternative measures.

[17]For example, using this method, an author named "Sarah" is assigned a 99.6 percent chance of being female, "John" a 0.01 percent chance, and "Francis" a 51.5 percent chance.

[18]We attempted a similar exercise with the first and last name of Black inventors. This approach led to many potential false positives, because there are fewer distinctively Black names in the historical censuses at this time. Given the low rate of patenting among Blacks during this time period, these false positives precluded a similar analysis for Blacks.

keywords and fine-tune the parsing rules to minimize false positives. This process is repeated until we are unable to identify new keywords. We report the final list of keywords and associated parsing rules in Appendix B.

Last, we are interested in identifying first-time inventors. We do so by clustering inventors with the same name, using their reported location to disambiguate them. More precisely, we assign two inventors on separate patents the same ID if their first names start with the same letter, more than 90 percent of their full names match,[19] their residence is within a 50km radius, and they patented within 10 years of one another.[20] While this method is not perfect (e.g., it will create two separate clusters if an inventor moves across the country), it strikes a good balance between precision and recall. After assigning unique IDs to each inventor, we identify each inventor's first patent.

3.3 City and county covariates

We construct city and county-level covariates from historical census data and related sources. We use each city's time-varying population collected by Erik Steiner and Jason Heppler.[21] For other standard covariates, including sex shares, race shares, average age, share of the population enrolled in school, and the occupation and industry of workers, we use the 1900 census micro-data aggregated at the city level. We calculate the fraction of each city's laborers in the mining industry in 1900 using census occupation responses. We also examine other details of the occupation structure in each city. In particular, we calculate the proportion of workers in "innovation-focused" occupations as defined by first-digit 1950 occupation codes: professionals (including engineers), managers, skilled craftsmen, and skilled mechanical operators. To calculate a proxy for city-level earnings and aggregate the entire industry and occupation distribution, we apply Saavedra and Twinam (2020)'s predicted earnings algorithm based on state, sex, age, race, occupation, and industry.[22]

For robustness, we use two county-level proxies for union activity measured before 1900. First,

[19]We use a fuzzy matching strategy because the names are extracted from digitized documents and sometimes contain typos due to OCR errors. The actual procedure takes into consideration the length of both strings and corresponds to the percentage of characters that match when the two names are of same length. We provide more details in Appendix B.

[20]We also allow for a larger radius (400km) if the two patents were filed within 5 years of one another.

[21]See https://github.com/cestastanford/historical-us-city-populations/ for a full description of the data. A number of cities do not have consistent population data in this source. We augment this series from records in the 1900 census and scraped data from Wikipedia to create a population estimate at the time of library entry for every city in the sample. In the regressions, we interpolate population between decennial censuses.

[22]Censuses before 1940 did not ask for income information.

we use the count of Knights of Labor assemblies formed before 1900. The Knights of Labor were the first sizable, national U.S. union. They were founded in 1869 and at their 1886 peak represented 20 percent of industrial laborers (Bittarello, 2019). The county-level Knights of Labor data were originally digitized by Garlock (1982) from contemporaneous union press and organizational publications. We use an updated version of Garlock's data compiled by Bittarello (2019).[23] Second, we use information on the number of strikers involved in labor disputes from 1881 to 1894. These data were originally compiled in the Third and Tenth Reports of the Commissioner of Labor in 1888 and 1896.[24] We use a geocoded version of these data recently compiled by Bittarello (2019).

Finally, we digitize data on city-level historical college locations from the US Bureau of Education's 1902 *Report of the Commissioner of Education*. This document records information on all colleges as they existed in the academic year 1900-1901. We collect the locations of all colleges and universities that grant A.B., B.S., or B.L. degrees, including women's colleges and technical universities.[25]

3.4 Summary statistics

Figure 3 shows the cumulative distribution of grant dates for cities that built (panel A) and rejected (panel B) Carnegie libraries. Both panels display similar shapes. Very few libraries were sponsored until the late 1890s. Starting in 1900, the number of granted libraries sharply increased, before settling into a steady trend after 1905 and dramatically slowing in the late 1910s.

Figure 4 shows the geographic distribution of Carnegie libraries across time and locations. Each panel plots the locations of Carnegie libraries that had been granted by the indicated date. The lowest figure (panel E) maps the location of all Carnegie libraries. As this map illustrates, the reach of Carnegie's program was national. Almost every state received at least one Carnegie library. Despite this outreach, some geographical patterns are clear. Granted libraries were popular in the Midwest and Northern states.[26] Indiana had the most granted Carnegie libraries, with a total of 164

[23]As in Bittarello (2019), we use the count of Knights of Labor assemblies and not membership because the membership data is spotty in later years. Moreover, branches had an incentive to exaggerate membership to the national organization to secure additional funding.

[24]These reports are widely used in studies of the early labor movement, including Card and Olson (1995), Rosenbloom (1998), and Currie and Ferrie (2000).

[25]Technical universities in the 1902 report include institutions like MIT, as well as public universities like Texas A&M and Purdue. The specific colleges and universities that we use are found in Tables 29, 32, and 36 of the 1902 report.

[26]While Southern states had fewer people, this result also holds on a per capita basis.

libraries built in 157 cities.

Figure 5 shows a similar map with the addition of rejecting cities as red, larger dots. As with cities that received a library, rejecting cities are located all across the United States. Cities in Southern states were more likely to reject libraries conditional on receiving an offer. For that reason, throughout the rest of the paper, we confine most of our discussion to *within*-state comparisons of accepting and rejecting cities.[27]

Figure 6 compares covariates across cities that received and those that rejected Carnegie libraries measured before library receipt. We plot the coefficient from an indicator variable for building (rather than rejecting) a Carnegie library from a regression on standardized covariates conditional on state fixed effects. The covariates considered here include population, share of women, average age, share of Blacks, average predicted earnings based on occupation and industry, share of the population currently enrolled in school, share of the workers in the mining industry and in innovation-focused occupations, and the number of Knights of Labor Assemblies and strikers observed within five miles before 1900.[28] Our identification strategy does not require balance on pre-treatment characteristics, but these results suggest that, conditional on state, places that accepted and rejected Carnegie libraries were broadly similar on many observable characteristics before Carnegie's program. In our empirical analysis, we show estimates from regressions when we do and do not flexibly control for these covariates.

By contrast, Figure 7 shows a starkly different result when we make within-state comparisons between cities that built Carnegie libraries and those that *did not apply* for a grant. We use the same variables and comparison strategy described in the previous paragraph. While cities that apply look similar to each other, Figure 7 illustrates that cities that did not apply are systematically different. In particular, non-applicants have fewer workers in plausibly high-innovation occupations—particularly craftsmen—and have lower imputed earnings. We do not use these non-applicant cities in our baseline analysis, although we do use them to estimate the possible spillover effects of libraries.

[27]We also show that our results are robust to excluding the South from our analysis.

[28]Each covariate is standardized to have mean zero and standard deviation 1 so that it is possible to plot them on the same scale. The standardization does not affect the interpretation of the results.

4 Empirical analysis

In this section, we describe our identification strategy and present our main results. First, we show that the raw data suggests an impact of libraries on patenting. Next, we describe our regression-based strategy, situate our approach relative to recent methodological advances in difference-in-differences designs, and present the main patenting results of the paper.

4.1 Patenting trends

In Figure 8, we plot city-level patenting for our treatment and control samples twenty years before and thirty years after library grants. We show the log number of patents conditional on grant year fixed effects over time. Cities that built and rejected Carnegie libraries follow parallel patenting trends before libraries were granted. The trends start diverging shortly after grants, consistent with construction times. The dashed red line indicates the average number of years in our sample from the moment the grant was received to the opening of the library. Patenting differences between cities that accepted and rejected libraries peak between 5 and 15 years after library receipt.[29] This difference declines over time, and by 30 years after library acceptance cities that accepted and rejected libraries revert to similar levels of patenting.[30]

4.2 Difference-in-differences estimates

To formalize the patterns that we observe in the raw data, we estimate difference-in-differences regression models of the form:

$$PatentMeasure_{i,s,t} = \beta_1 Post_{i,t} + \beta_2 Library_i \times Post_{i,t} + \delta_{s,t} + \gamma_i + \epsilon_{i,s,t} \quad (1)$$

where $PatentMeasure_{i,s,t}$ is a measure of patents in city i, state s, year t. $Library_i$ indicates cities that constructed a Carnegie library. $Post_{i,t}$ is a dummy variable that takes value 1 in the years after city i received a library grant; it is a function of both time (t) and the grant year of city i. $Post_{i,t}$

[29]If we plot this figure in levels instead of logs, the treatment effect grows more gradually over time, as demonstrated in Figure A6. The pattern of treatment effect growth is consistent with the timing of results that Furman et al. (2021) find for the effects of patent deposit libraries.

[30]This is likely due to an increase in the complexity of inventions over time that reduced the importance of public libraries in promoting innovative activities.

is well-defined for all units in both our treatment and control group since all cities were offered library grants.[31] The baseline empirical model includes state-by-year ($\delta_{s,t}$) and city (γ_i) fixed effects. For our main analysis, we use $ln(patents + 1)$ as our patent measure to make magnitudes directly comparable to other recent papers; we later show that our results are robust to other dependent variable transformations.[32] The coefficient of interest, β_2, identifies the average increase in patenting after receiving a grant in cities that built a library relative to cities that were offered a grant but did not build one. We estimate our model using observations 20 years before and after library grants for each city.[33] Unless otherwise indicated, we cluster standard errors at the city level.

Our specification of Equation 1 relates to concerns in the applied econometrics literature about the interpretation of staggered difference-in-differences estimators. When units are treated at different times, post-treatment periods for never treated units are not usually defined. Instead, researchers often estimate models with unit and time fixed effects and a time-varying treatment indicator that is always zero for never-treated units, colloquially called the two-way fixed effect (TWFE) approach.[34] Our specification departs from these models as we observe grant dates for all cities. Equation 1 includes well-defined post-period indicators ($Post_{i,t}$) for both treated and control units. This specification is equivalent to making a simple comparison: we calculate patenting changes before and after grants in cities that built their libraries relative to cities that did not, conditional on included covariates and fixed effects. Moreover, we are able to specify models that only exploit variation in treatment status among groups of cities that received library grants in the same year by including grant year and grant-by-calendar year fixed effects.

To interpret β_2 as the causal effect of libraries, the standard difference-in-difference assumption

[31]We use grant dates and not library construction dates because grant dates are well-defined for all units in our sample, while library construction dates are not. Moreover, the gap between when a library is granted and opened may be a function of other city characteristics that are correlated with patenting.

[32]We establish this systematically in Section 6, where we show results from estimation using the inverse hyperbolic sine of patents, patent counts, Poisson-modelled patent counts, zero-inflated Poisson-modelled patent counts, and aggregated versions of these measures where we collapse each city into a before and after library grant observation.

[33]This implies that each geographical unit in our baseline sample has the same number of observations before and after library entry. In subsequent analysis, we estimate dynamic models that allow the effects of Carnegie libraries to differ over time. In Section 6, we show that our results are robust to alternative choices for the length of the pre-period.

[34]A recent and rapidly expanding literature examines the properties of the TWFE estimator. For example, Goodman-Bacon (2021) and De Chaisemartin and d'Haultfoeuille (2020) show that the TWFE method may estimate a biased parameter. In particular, the TWFE approach in part uses units treated in earlier time periods as controls for later treated units. If treatment effects differ over time, these comparisons will be biased, affecting and potentially even reversing the sign of the aggregated TWFE parameter (Goodman-Bacon, 2021). These issues are not limited to aggregated models—they also affect event study TWFE approaches that allow estimated treatment effects to differ relative to the treatment period (Callaway and Sant'Anna, 2020).

must hold: conditional on included covariates, outcomes in cities that did not build their granted library *would have* followed the same path as cities that did build their libraries. While fundamentally untestable, we demonstrate the plausibility of this assumption by showing that pre-grant trends in patenting are similar across our treatment and control groups and that our results are stable with and without a wide range of potential controls and fixed effects.

Our main results, obtained by estimating Equation 1, are presented in Table 1. We show eight different specifications, each corresponding to different combinations of fixed effects and included covariates. Estimates for the key coefficient (Built library $\times$ post) suggest that the number of patents increased by approximately 7-11 percent in the years after the receipt of library grants in cities that built their granted libraries relative to cities that did not build libraries. The second panel of Table 1 presents results for the same specifications where we exclude observations after 1928, to avoid using observations that overlap with the Great Depression and subsequent recovery. Effects are larger (8-14 percent) when we exclude the Great Depression, which led to a sharp decrease in national patenting.

We also estimate dynamic versions of Equation 1 that allow the impact of libraries on patents to vary over time. In particular, we estimate versions of

$$PatentMeasure_{i,s,t} = \sum_{r=-20}^{30} [\beta_1 \times RelYear_{i,t,r} + \beta_{2,r} Library_i \times RelYear_{i,t,r}] + \delta_{s,t} + \gamma_i + \epsilon_{i,s,t} \quad (2)$$

where the terms in Equation 2 are defined similarly as in Equation 1. However, instead of a single post period indicator, we interact treatment status with a set of dummy variables for years relative to Carnegie grants ($RelYear_{i,t,r}$). We bin years in five-year increments to maximize power while allowing for a flexible estimation of treatment effects over time. The vector of coefficients $\beta_{2,r}$ traces out how the relative patenting differences between cities that did and did not build granted Carnegie libraries change over time. Because we estimate these patenting differences both before and after the date of treatment, this specification can also be used to assess the plausibility of our empirical strategy. If the estimated impact of libraries on patenting were trending upward even before libraries were built, this would imply that accepting cities were positively selected on patenting trends and that post-library patenting differences are unlikely to reflect the causal impact of libraries.

Figure 9 plots the marginal effects of receiving rather than rejecting a library evaluated at each bin relative to the grant year. The excluded category is 1 to 5 years before library grant dates. The reported coefficients reflect five-year bins containing the labelled relative year and the four following years.[35] Standard errors are clustered at the city level and we plot 90 percent confidence intervals. The results in Figure 9 are consistent with the dynamics outlined by the raw data plotted in Figure 8. Flat pre-trends indicate that cities that received and rejected libraries follow similar patenting patterns before library grants. Quickly after library grant receipt, and consistent with data on construction time, patenting behavior diverges. Differences peak between 5 and 15 years after library entry before they start to converge. Figure 10 shows that these results are broadly similar if instead we estimate the model with either (a) city and grant by calendar year fixed effects, or (b) city, grant by calendar year, and state by calendar year fixed effects. In each case, we see generally flat pre-trends and a sharp increase in patenting after the libraries are built.

Taken together, these results suggest that Carnegie libraries increased innovation, measured through patenting activity, but that these increases were not permanent. This is consistent with the nature of innovation described in previous sections. While a 1910 inventor usually worked alone on more practical inventions, by the 1930s and 1940s patenting had become more technical and professionalized. This is illustrated in Figure 2, which shows a decreasing share of patents in classes like Human Necessities and Constructions over time, with technology classes like Chemistry and Physics increasing in their place. Patenting in more sophisticated technology classes likely required access to resources and lab space beyond the scope of public libraries.

4.3 Patent quality

Not all patents have the same innovative content, and even among the most innovative patents, real-life impacts and value can differ significantly. If libraries only increased lower quality patenting, the impacts discussed in the previous subsection would be overstated. To explore this possibility, we estimate the effect of library on multiple measures of patent quality. We estimate analogues of Equation 1 with a measure of patent quality in city i, state s, and year t as outcome variables:

$$QualityMeasure_{i,t} = \beta_1 Post_{i,t} + \beta_2 Library_i \times Post_{i,t} + \delta_{s,t} + \gamma_i + \epsilon_{i,s,t} \tag{3}$$

Table 2 includes the parameter estimates from this regression for three measures of quality: (1)

[35] For example, the bin labeled 5 contains the 5th-9th years after library grant dates.

the probability that a city-year observation had a patent that was cited in the future, (2) the count of future citations, (3) the probability that a city-year produced a patent in the top ten percentile of the Kelly et al. (2021) quality measure described in Section 3. Citations measure whether a patent had lasting relevance to future inventions, while the Kelly et al. (2021) text-based approach measures whether a patent is both novel for its time and inspires future innovation.

Table 2 shows that libraries increased both the probability of observing a patent that later garnered citations and the number of citations that these patents received. We observe a 4-7 percentage point increase in the probability of observing a cited patent (relative to a mean probability of 0.304), and the number of citations increases by 1–2 patents (relative to a mean of 2.83). However, we find no effect on the probability of producing a top 10 percentile patent using the Kelly et al. (2021) measure; estimated coefficients are centered around zero and precisely estimated, ruling out large positive or negative changes. These findings suggest that the results described in the preceding subsection are not driven by changes in patent quality. Libraries appear to have increased the patenting rate without a decrease in quality. If anything, they slightly increased patent quality—as measured by citations.

4.4 Spillovers to nearby cities

In the analysis so far, we have compared patenting across cities that built and rejected Carnegie libraries. Our estimates will differ from the aggregate effect of libraries if there are spillovers to nearby towns.[36] These spillovers could make us overstate or understate the true causal effect of libraries. For example, if libraries attracted users from outside the city, they could have had a direct impact on innovation in nearby areas, and our estimates would understate the actual impact of libraries.[37] On the contrary, if would-be innovators move to or file patents in cities with libraries that they would have otherwise filed in their hometown our estimates would be overstated.

To test for spillover effects, we construct a set of 'doughnut' treatment and control groups incorporating cities at increasing distances from accepted and rejected Carnegie libraries. We define cities within 15, 30, and 45 miles of Carnegie libraries as separate treatment groups, *excluding* the

[36] Butts (2021) provides a useful methodological discussion of the potential issues in difference-in-differences models when spillovers are present.

[37] Indeed, historical records suggest that cities took into account the size of neighboring populations when applying for libraries and, when negotiating grant amounts, a number of cities argued that they needed more money because their library would draw users from beyond their city borders (Bobinski, 1969).

Carnegie library city itself from all samples. We similarly define cities within 15, 30, and 45 miles of a rejecting library (excluding the rejecting city itself) and *not* within those distances of a Carnegie library city as corresponding control groups. We re-estimate Equation 1 using the new spatial treatment and control samples separately for each distance threshold. We assign grant dates based on the closest accepted or rejected library and cluster standard errors at the city-of-grant level.

Figure 11 shows the result of this exercise when we estimate the model with city and state-year fixed effects. The leftmost estimate is the baseline difference-in-differences estimate from Equation 1. This estimate reflects a comparison of Carnegie and rejecting cities before and after grant dates and is identical to the corresponding estimate reported in Table 1. The remaining estimates are for identically specified models with increasingly dispersed treatment and control groups, as defined above. The results show that there is evidence of spillover effects within 15 miles. These effects quickly decrease to zero as distance increases. The results in Figure 11 use nested distance treatment definitions; alternatively, we can define non-nested treatment bins—for example, instead of estimating a coefficient for 0-30 miles we estimate separate coefficients for 0-15 and 15-30 miles. Results using non-nested bins are shown in Figure A9. The figure shows that the declining treatment effect patterns seen in Figure 11 are the result of positive spillover effects in the 0-to-15 mile distance bin and null effects when considering towns located father apart. This finding is consistent with prior work on patent spillovers, which finds that they are geographically concentrated (Jaffe et al., 1993; Murata et al., 2014). Thus, our baseline findings do not appear to be overstated due to "brain-drain" from nearby areas. Instead, it appears that libraries have modest but positive spillover effects on nearby communities, consistent with the recognition from contemporary library leaders that their new buildings would draw users from nearby areas. Moreover, these results suggest that the effects that we are observe are not due to differences in the broad regions where Carnegie libraries were built: effects are null outside 15 miles.

5 Potential mechanisms and heterogeneity

In this section, we consider two possible mechanisms that could explain the patenting increases that we observe: (1) Libraries may have increased the stock of available knowledge in cities via books, magazines, and other materials and (2) Libraries may have increased collaborative opportunities between inventors. We also investigate whether certain demographic groups or cities were more likely to be affected by the libraries.

5.1 Mechanisms

Access to library information might not affect all types of patenting behavior equally across technological classes. In fact, the ALA's 1904 book guide for new libraries included many practical "how-to" books for trades and agriculture. Intuitively, it seems plausible that public libraries would have a smaller impact on the most technical inventions that require a large amount of human (and possibly physical) capital. Even in the early 1900s, the most technical scientific material was likely to be only available in research libraries and at universities. Moreover, the most technical fields also often required access to expensive machinery outside the reach of many citizens. To study the differential effect of access to libraries on patenting behavior, we separately estimate Equation 1 for each of the eight main subgroups in the Cooperative Patent Classification (CPC), which identify the primary industrial applications of each patent.

Consistent with library holdings, the estimates reported in Table 3 suggest that libraries had the largest effects on patenting in classes that correspond to the practical trades, such as construction, transport, and mechanical engineering. These heterogeneous effects are in line with historical records of the books that libraries commonly carried. Patenting in topics that overlap with library collections increased proportionately more in cities that accepted Carnegie libraries. By contrast, we estimate no effects in Chemistry and Physics: coefficients are small and statistically indistinguishable from zero.

We also test whether patent citations to prior materials increased after Carnegie library construction. If patrons effectively used materials in libraries for their inventions, we might expect the number of these citations to increase. We select patents that cite prior materials by identifying an initial set of words that are associated with citation in a training dataset and searching the corpus of remaining patents for similar phrases.[38]

We estimate analogous models of our baseline difference-in-differences regression from Equation 1. For this analysis, the outcome variable is an indicator for whether a patent in a particular city-year cited a book, previous patent, or magazine. Our results, in Table 4, indicate that patents that cite past materials increased more in cities that built rather than did not build Carnegie librar-

[38]More details on this procedure are given in Appendix B.

ies. In particular, building a library is associated with approximately a 0.5 to 0.7 percentage point increase in the probability of observing a book-citing patent, though the effects are imprecisely estimated in some samples. The largest and most precisely estimated effects arise when we focus on the pre-1925 sample.[39] This effect is large relative to the baseline probability of observing prior-work-citing patents (1.2 percent). Table 4 also shows that the count of patents that cite prior work increases after libraries are built.

Knowledge is not only available in books and other physical media. It can also be accessed through collaboration with other people. Such exchanges may increase innovative output—for example, Andrews (2020) finds that a reduction in collaboration after prohibition reduced aggregate patenting. For this reason, we examine whether libraries affected collaboration. Community-centric programming was common at libraries in the early 20th century. Carnegie's suggested library blueprints included meeting rooms for community activities, and contemporaneous accounts of the Carnegie program note the variety and number of groups that met at libraries (Learned, 1924).

To estimate the effect of libraries on collaboration, we consider the number of multi-authored patents per city-year. If collaboration became more common after library entry, we would expect the number of multi-authored patents to increase more in cities that received a Carnegie library relative to those that did not build a library. We estimate analogues of our baseline difference-in-differences regression with an indicator for observing a multi-authored patent as the outcome variable. As with the book citation mechanism results discussed above, we estimate these effects on multiple samples, focusing our attention on short-run library effects.

Table 5 shows results from this analysis. These estimate indicate that the probability of observing multi-authored patents generally increases in cities that built Carnegie libraries, though the results are often imprecisely estimated. For example, in the third row we observe an imprecisely estimated increase in the probability of observing a multi-authored patent of 2.6 percentage points in the pre-1925 sample, relative to a baseline probability of observing a multi-authored patent of 18 percent. Table 5 also shows that the count of multi-author patents increases after libraries are built.

[39]In the table we present results for the full, pre-1929, and pre-1925 sample. Because the patenting results are concentrated in earlier years, restricting our sample allows us to better identify the effect of libraries on rare outcomes, such as patents that cite prior materials.

5.2 Heterogeneity

Since libraries were open to all, they might have been particularly useful for those who did not have as much access to other sources of information and education at this time in history. This includes women and immigrants, both of whom were underrepresented in patenting. Indeed, librarians were one of the first professionalized career paths open to women, along with nursing and teaching. Moreover, libraries might have been particularly important for new inventors, since those with more experience might already have had access to alternative sources of knowledge. By contrast, the qualitative evidence that we have reviewed suggests that established inventors also heavily used libraries. In this subsection, we analyze these and additional dimensions of heterogeneity.

First, we test whether libraries affected the intensive or extensive margin of patenting. In Table 6, we show similar specifications as our main table, with an indicator for whether a city-year had a patent as the outcome variable. We see positive impacts of libraries on the probability of observing a patent, with generally larger effects when we focus on the pre-1929 sample. However, these effects are small, less precisely estimated than our main results, and differ across specifications, with point estimates ranging from 0.4 to 3 percentage points. While libraries may have affected the probability that a patent is observed in a particular city, our results seem to be mainly driven by increases in cities that already had some patenting activity.

In Table 7, we perform a similar exercise, except that we look at the number of *first-time* inventors who filed a patent in a given year and the share of first-time inventors as a fraction of total patents filed in a given city-year. The results in Table 7 suggest that while the total number of first-time inventors per city-year increased by 0.2–0.4 people after libraries were built (from a mean of 1.35), their overall share did not change much, as established inventors were also increasing their productivity. These results provide little evidence that libraries differentially affected either established or novice inventors. Instead, it appears that the two groups of inventors were similarly affected.

Next, we estimate the effect of Carnegie libraries on women and immigrants using a version of Equation 1 with each subgroup's patent counts as the outcome variable. We infer gender and immigrant status using names that we extract from each patent, as described in Section 3, and aggregate the count of women and immigrant patents to the city-year level. Table 8 presents the results from this analysis. Panel A shows the results when the count of women patenting is the dependent variable; Panel B shows analogous results for immigrants. The results consistently indicate that patenting increased for both women (by 0.02–0.04 patents per-city year) and immigrants (by 0.09–0.14 patents per-city year) after receiving a library grants. We also estimate the effect of libraries on the share of women and immigrant patents. As with first-time inventors, we see little change in these shares, likely because libraries also impacted non-women and non-immigrants.

Beyond inventor characteristics, we also explore the characteristics of *cities* that might have changed how effective libraries were at increasing innovation. We estimate versions of Equation 1 that interact the key difference-in-differences coefficient with a given city characteristic variable.[40] In particular, we separately interact the key difference-in-differences coefficient with an indicator for whether a town had a college and whether it was in the top half of the sample distribution of: youth in school, imputed income, share craftsmen, share Black, and population. A positive coefficient in this table indicates that libraries were more effective at increasing patenting when the indicated condition is satisfied. For example, a positive coefficient in the "had college" row would indicate that Carnegie libraries had a larger effect on innovation in cities that had colleges.

The estimates are presented in Table 9. Overall, we find imprecisely estimated but positive complementarity between libraries and the share of youth in school, the share of craftsmen in the labor force, and the size of the town. While the limited sample size does not give enough statistical power to detect small heterogeneous effects, our results suggest that the effects of libraries did not dramatically differ across the city characteristics considered here.

These findings relate to the external validity of our results. Because our estimates are generated from a comparison among cities that applied for Carnegie libraries, we cannot necessarily extrapolate our findings to predict what would have happened had non-applicant cities received libraries. However, Carnegie library recipients were a diverse group, ranging from growing cities like Columbus, Ohio to small rural enclaves like Bayliss, California. The limited evidence of heterogeneity in patenting effects across multiple dimensions of city characteristics described in this subsection suggests that our results are not limited to specific types of cities. Importantly, libraries did not require an already-educated population or a college to be effective—they were impactful even when the baseline human capital stock or population were lower, consistent with Carnegie's claim that libraries were "palaces for the people."

[40]We fully saturate the interactions, so these models also separately include Built library × covariate and post × covariate variables.

6 Robustness of patenting results

In this section we describe a set of exercises that establish the robustness of our main results. These robustness checks fall into two broad categories: alternative sample selection criteria and alternative patent measure/specification choices.

Table 10 shows a collection of the key sample robustness checks. The first panel shows results from Equation 1 for subsets of the baseline sample. First, since the South was slower to build public good infrastructure during this time period, we show results excluding all Southern states. Next, we show results excluding Carnegie's two "home" states for himself and his businesses, respectively New York and Pennsylvania. Third, we estimate our baseline model using only 15 or 10 years of pre-period observations, instead of the 20 years used in our main analysis. Fourth, we exclude a number of control towns whose stated primary reason for rejecting the grant appears to be financial according to Bobinski (1969). In particular, these cities may have been worried about Carnegie's requirement that cities would pledge 10 percent of the cost of the building on an annual basis for maintenance. Finally, we in turn exclude all cities larger than 15,000 and 5,000 people. These two restrictions test whether the innovation-promoting effects of Carnegie libraries also operated in small towns and rural areas in addition to larger cities. Estimated coefficients are modestly smaller (0.079 compared to a baseline estimate of 0.097) when we focus on cities with less than 5,000 people, consistent with the heterogeneity results described in the previous subsection.

The second panel of Table 10 shows a similar set of exercises where we instead add cities to the baseline sample. We re-estimate Equation 1 after adding pre-1899 grant cities, high population cities and counties, control cities that built libraries from local philanthropists, and cities missing 1900 covariates. In the last row of the table, we estimate the baseline model after relaxing *all* sample selection criteria. In all cases, results are similar to the baseline estimates.

In Table 11 we show a series of robustness checks using alternative specifications and patent measures. First, we show a model that conditions on time-varying log city population.[41] Andrews (2021b) finds that population growth could explain almost all of the effects of colleges on local patenting. By contrast, the results reported in Table 11 suggest that, in the case of libraries, population growth is not responsible for our results. Next, we show estimates from models that condition on

[41] We use a subset of cities that have valid population data for this analysis.

pre-Carnegie measures of county-level union activity that we described in Section 3. In particular, we condition on the pre-1900 log number of Knights of Labor Assemblies interacted with the post dummy, the pre-1900 log number of strikers interacted with the post dummy, and both measures together. We see almost no change in our estimated difference-in-differences coefficients, implying that our results are not driven by differences in union behavior that might be correlated with future patenting. Last, we show results after conditioning on whether a city had a college or university interacted with the post dummy and after conditioning on the interaction of the share of children in school in 1900 and the post dummy. In both cases, we see almost no change in our estimated Library $\times$ Post coefficient, suggesting that our results are not merely picking up the time-varying effects of pre-existing levels of human capital on innovation.

Next, we show results using two alternative transformations of the number of patents as outcome variables, the inverse hyperbolic sine of patents and untransformed patent counts. Table 11 shows that when we use these patent measures, the results are positive and in percentage terms consistent with or larger than our baseline estimates.[42]

In Table 11 we also show results from three maximum likelihood models: the Poisson count model, the zero-inflated Poisson count model, and the Negative Binomial model.[43] For each, we report the mean marginal effect (in patent counts) of the key difference-in-differences coefficient. In each case, results are comparable or larger in percentage terms relative to our baseline log estimates.

To provide further evidence that our results are similar across patent transformations, we estimate aggregate versions of our baseline regression models. In particular, we sum the number of patents for each city before and after library grants are made. Thus, each city has two observations. We then take log and inverse hyperbolic sine transformations, as well as estimate the Poisson model described above. Since we cannot estimate state-year fixed effects in these specifications, we report models with city fixed effects and city plus state-post period fixed effects. Results are shown in Table A4, and are similar or larger than our baseline estimates.

We demonstrate that our event study results are also robust to alternative specifications. Figures 10 and A2 shows that our baseline event study model looks very similar if we use models with

[42]Tables A2 and A3 report the full set of fixed effects specifications for the inverse hyperbolic sine transformation and untransformed patent counts, respectively. These tables show a similar pattern of results as our main $ln(pat + 1)$ specifications.

[43]For the zero-inflated Poisson, we use city population as a predictor of 0-patent status in the secondary estimation equation.

alternative sets of fixed effects. Similarly, Figures A3, A4, and A5 show that across fixed effect specifications, our event studies look similar if we use inverse hyperbolic sine transformations of our patent variable instead of the $ln(pat + 1)$ transformation that we use for our baseline analysis. Similarly, Figure A6 shows that the raw trends in patenting across cities look similar if we use the patent count outcome instead of logged patents. Figure A7 shows a version of our event study analysis that uses yearly instead of binned data. The result is similar to our main result, though standard errors for point estimates are larger. Figure A8 shows a version of the yearly results where we use the library opening date (instead of grant dates) as our defined treatment start period—our results are similar.[44]

Last, Figures A10 and A11 shows that the spillover analysis is also robust to changes to the main model specification. Figure A10 shows the baseline spillover estimate when using city and year fixed effects. Figure A11 shows the same estimates when we estimate each distance bin separately as opposed to using nested bins. We find very similar patterns using this alternative specification.

7 Conclusion

In this paper, we study the rollout of one of the most common institutions in local communities: the public library. Leveraging the expansion in library services generated by Andrew Carnegie's grants in the late 19th and early 20th centuries, we test whether cities that built libraries increased innovative activity. We find that patenting increased by 7-11 percent in cities that built libraries relative to a new control group of cities that did not build a library despite being deemed eligible to receive a grant. Effects peak between 5-15 years after library entry, before converging to zero as patenting became more complex and team-based. Libraries lead to modest patenting spillovers to nearby towns, suggesting that they were drawing users from outside of their immediate city borders. We show that our results are not driven by increases in low-quality patenting and are robust to a wide range of sample, measurement, and estimation choices.

To explain these results, we explore two potential mechanisms. We find that increased access to information and collaboration are potential explanations for our results. In particular, we see increases in the number of patents that cite prior work after libraries are built and our largest effects

[44]Because library opening dates are not observed for cities that did not build a Carnegie library, we assign the opening date for those cities as two years after their grant date—the median time to construction in the Carnegie library sample.

occur in practical patenting classes that overlap with the types of materials that libraries held. On the other hand, we see little effects of libraries in more technical fields where public library knowledge is less likely to be useful. We also find suggestive evidence that multi-author patenting increased after libraries entered, consistent with libraries serving as gathering places for the community.

Our setting provides a unique opportunity to test the hypothesis that information-providing institutions impact innovation. Unlike many of the technologies and institutions previously studied in this literature (e.g., the printing press; national science academies), the public library is a distinctly local institution focused on broadening knowledge access for everyone. Our results show that such institutions can have meaningful impacts on innovation. The geographic spread of public libraries provides a useful historical laboratory to study these channels, but libraries are by no means unique. While libraries today are unlikely to have the same impact on innovation as they did in the early 20th century, the spread of technical information via the internet and sources like Wikipedia and Google Patents likely play analogous roles worthy of further study (e.g., Thompson and Hanley, 2018; Derksen et al., 2019).

Our results also motivate a need for more research on the historical and contemporary effects of libraries. While the unique informational role of public libraries has been diminished by the emergence of the internet, today's libraries are more community-focused than ever, with programs aimed at entrepreneurs, job-seekers, children, seniors, and many more. The rise of these programs likely reinforce and strengthen the social channel that we documented in the early 20th century. We hope this work motivates more studies of these library programs and their impacts on cities, both throughout history and today.

References

Aghion, P., L. Boustan, C. Hoxby, and J. Vandenbussche (2009). The Causal Impact of Education on Economic Growth: Evidence From U.S. Brookings Papers on Economic Activity 1, Brookings.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics 130*(4), 1781–1824.

Andrews, M. (2020). Bar talk: Informal social interactions, alcohol prohibition, and invention. Ssrn working paper, SSRN.

Andrews, M. (2021a). Historical Patent Data: A Practitioner's Guide. *Journal of Economics and Management Strategy*.

Andrews, M. (2021b). How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges.

Baldwin, N. (2001). *Edison: Inventing the Century.* The University of Chicago Press.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2018, 11). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *The Quarterly Journal of Economics 134*(2), 647–713.

Berkes, E. (2018). Comprehensive Universe of US Patents (CUSP): Data and Facts.

Biasi, B. and P. Moser (2021). Effects of Copyrights on Science: Evidence from the WWII Book Republication Program. *American Economic Journal: Microeconomics.*

Bittarello, L. (2019). Organizing Collective Action: Labor Strife in the US in the 1880s.

Bobinski, G. S. (1968). Carnegie Libraries: Their History and Impact on American Public Library Development. *ALA Bulletin 62*(11), 1361–1367.

Bobinski, G. S. (1969). *Carnegie Libraries: Their History and Impact on American Public Library Development.* American Library Association.

Bryan, K. A. and Y. Ozcan (2021, 11). The Impact of Open Access Mandates on Invention. *The Review of Economics and Statistics 103*(5), 954–967.

Butts, K. (2021). Differences-in-Differences with Spatial Spillover.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with Multiple Time Periods. *Journal of Econometrics.*

Card, D. and C. A. Olson (1995). Bargaining Power, Strike Durations, and Wage Outcomes: An Analysis of Strikes in the 1880s. *Journal of Labor Economics 13*(1), 32–61.

Cardona, M., T. Kretschmer, and T. Strobel (2013). Ict and productivity: conclusions from the empirical literature. *Information Economics and Policy 25*(3), 109–125.

Currie, J. and J. Ferrie (2000). The Law and Labor Strife in the United States, 1881-1894. *Journal of Economic History*, 42–66.

Czernich, N., O. Falck, T. Kretschmer, and L. Woessmann (2011). Broadband infrastructure and economic growth. *The Economic Journal 121*(552), 505–532.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–96.

Derksen, L., C. M. Leclerc, and P. C. Souza (2019). Searching for answers: The impact of student access to wikipedia.

Electrical Worker (1901). *Untitled article, September 1901.*

Furman, J. L. and M. J. MacGarvie (2007). Academic Science and the Birth of Industrial Research Laboratories in the US Pharmaceutical Industry. *Journal of Economic Behavior & Organization 63*(4), 756–776.

Furman, J. L., M. Nagler, and M. Watzinger (2021). Disclosure and subsequent innovation: Evidence from the patent depository library program. *American Economic Journal: Economic Policy 13*(4), 239–70.

Garlock, J. (1982). *Guide to the Local Assemblies of the Knights of Labor.* Greenwood Press Westport, CT.

Gilpin, G., E. Karger, and P. Nencka (2021). The Returns to Public Library Investment.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics.*

Graham, S. and D. Hegde (2015). Disclosing Patents' Secrets. *Science 347*(6219), 236–237.

Gross, D. P. (2019, February). The Consequences of Invention Secrecy: Evidence from the USPTO Patent Secrecy Program in World War II. Working Paper 25545, National Bureau of Economic Research.

Hausman, N. (2020). University innovation and local economic growth. *The Review of Economics and Statistics*, 1–46.

Hegde, D. and H. Luo (2018). Patent Publication and the Market for Ideas. *Management Science 64*(2), 652–672.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics 133*(2), 927–991.

Jaffe, A., M. Trajtenberg, and R. Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics 108*(3), 577–598.

Jones, T. (1997). *Carnegie Libraries Across America: A Public Legacy.* Wiley.

Kantor, S. and A. Whalley (2014). Knowledge spillovers from research universities: evidence from endowment value shocks. *Review of Economics and Statistics 96*(1), 171–188.

Kantor, S. and A. Whalley (2019). Research proximity and productivity: long-term evidence from agriculture. *Journal of Political Economy 127*(2), 819–854.

Karger, E. (2021). The Long-Run Effect of Public Libraries on Children: Evidence from the Early 1900s.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights, forthcoming*.

Kevane, M. J. and W. A. Sundstrom (2014). The Development of Public Libraries in the United States, 1870–1930: A Quantitative Assessment. *Information & Culture 49*(2), 117–144.

Kevane, M. J. and W. A. Sundstrom (2016a). Public Libraries and Political Participation, 1870-1940.

Kevane, M. J. and W. A. Sundstrom (2016b). State Promotion of Local Public Goods: The Case of Public Libraries, 1880-1929.

Krass, P. (2011). *Carnegie*. John Wiley & Sons.

Learned, W. S. (1924). *The American Public Library and the Diffusion of Knowledge*. New York: Harcourt, Brace.

Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press.

Moser, P. (2005). How do patent laws influence innovation? Evidence from nineteenth-century World's Fairs. *American Economic Review 95*(4), 1214–1236.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014, December). Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *The Review of Economics and Statistics 96*(5), 967–985.

Nasaw, D. (2007). *Andrew Carnegie*. Penguin.

Researching NYC (2015). Chester F. Carlson.

Rosenbloom, J. L. (1998). Strikebreaking and the Labor Market in the United States, 1881-1894. *Journal of Economic History*, 183–205.

Saavedra, M. and T. Twinam (2020). A Machine Learning Approach to Improving Occupational Income Scores. *Explorations in Economic History 75*, 101304.

Sharp, K. L. (1893). The ALA Library Exhibit at the World's Fair. *Library Journal 18*(8), 280–84.

The Coca-Cola Company (2015). The History of the Coca-Cola Contour Bottle.

The Los Angeles Times (2012, Oct). Stanford Ovshinsky Dies at 89; Inventor Founded New Field of Electronics.

Thompson, N. and D. Hanley (2018). Science is shaped by wikipedia: Evidence from a randomized control trial.

United Nations (2018). World's Most Vulnerable Countries on Track to Achieve Universal Internet Access by 2020 – UN report – United Nations Sustainable Development.

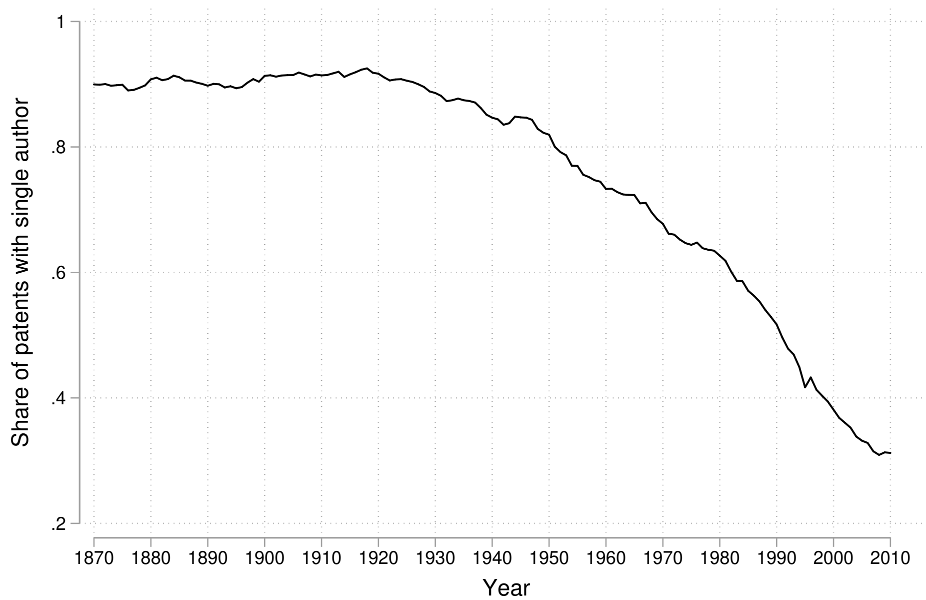
Weitzman, M. L. (1998). Recombinant Growth. *The Quarterly Journal of Economics 113*(2), 331–360.

Wenyon, M. (2009, Aug). NYPL, Mother of Invention.

Wiegand, W. A. (2015). *Part of Our Lives: A People's History of the American Public Library.* Oxford University Press.

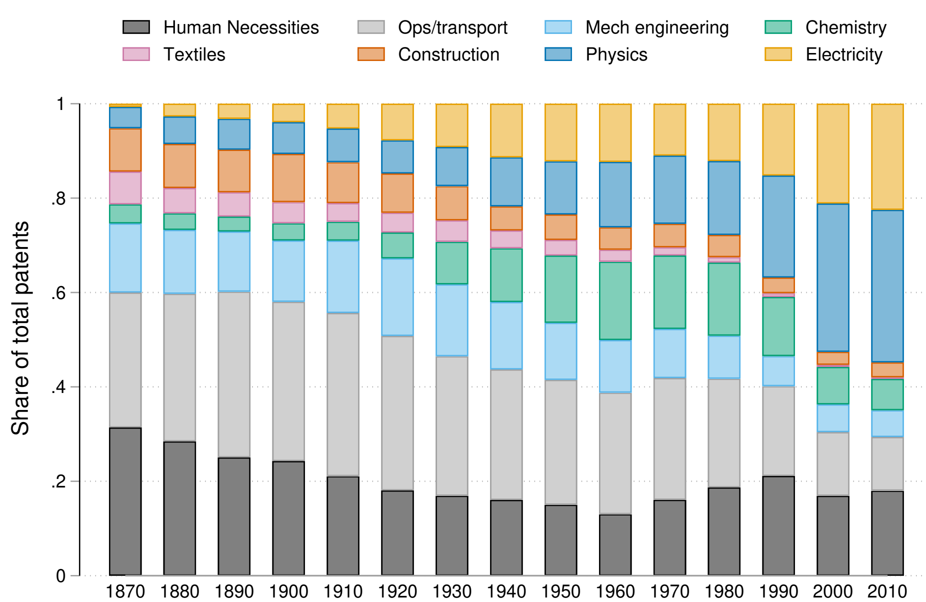
Xu, X., A. Watts, and M. Reed (2019). Does access to internet promote innovation? a look at the us broadband industry. *Growth and Change 50*(4), 1423–1440.

Figure 1: **Share of solo-authored U.S. patents by filing year**



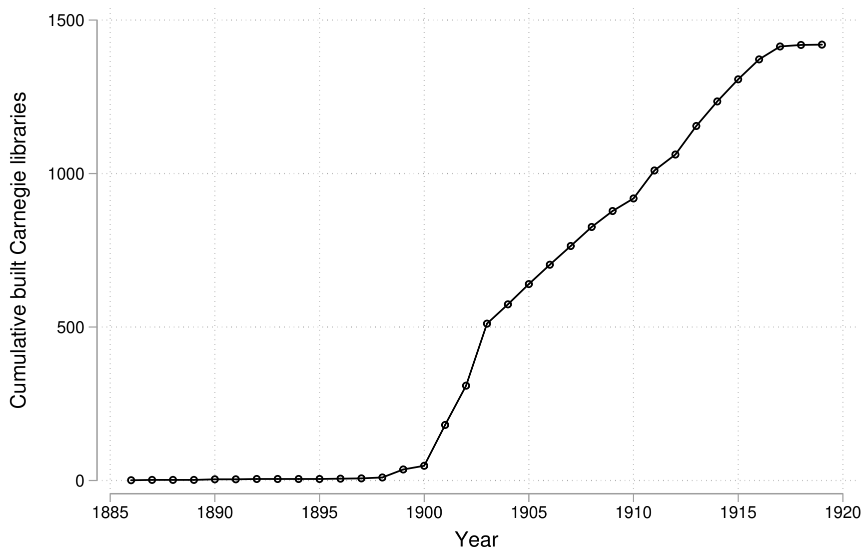
Notes: This figure shows the share of U.S. patents that had a single author, by filing year. Source: Comprehensive Universe of U.S. Patents (CUSP) data.

Figure 2: **Share of U.S. patents by Cooperative Patent Classification class by decade**

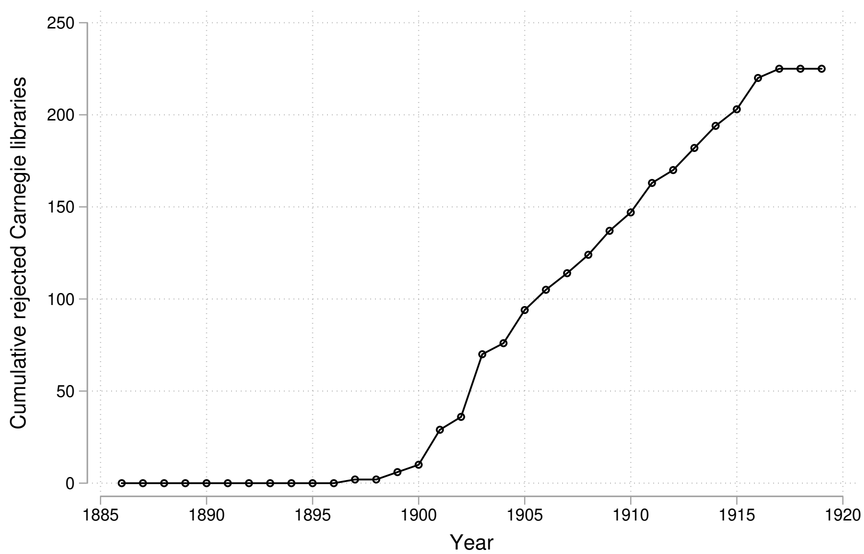


Notes: This figure shows the share of U.S. patents in each Cooperative Patent Classification grouping, by decade. Source: Comprehensive Universe of U.S. Patents (CUSP) data.

Figure 3: **Cumulative grant date distribution for Carnegie library and rejector cities**



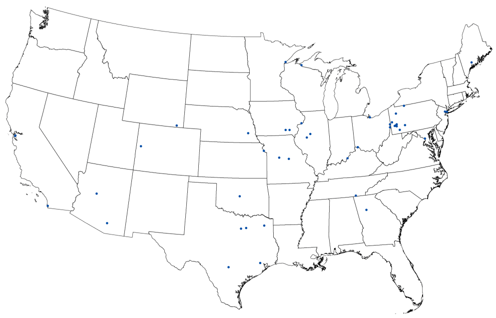
(a) Cities that built Carnegie libraries



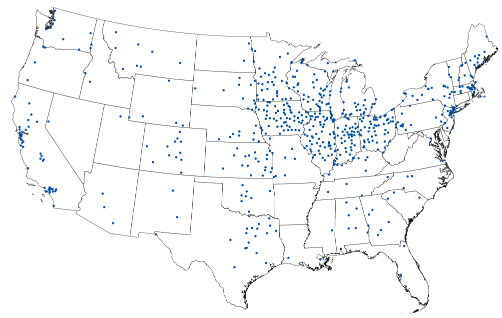
(b) Cities that rejected Carnegie libraries

Notes: These figures show the cumulative distribution of accepted (Panel a) and rejected (Panel b) library cities by year of Carnegie grant. For cities that received multiple library grants, the earliest grant year is used. Source: Bobinski (1969) and Jones (1997).

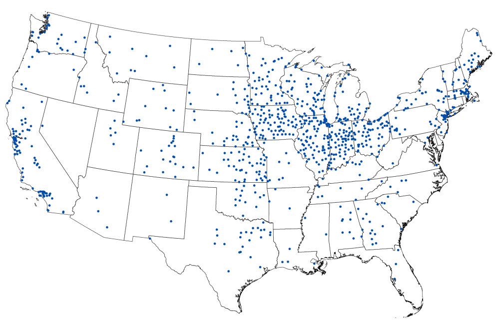
Figure 4: **Map of Carnegie libraries, by location and time of library grant**



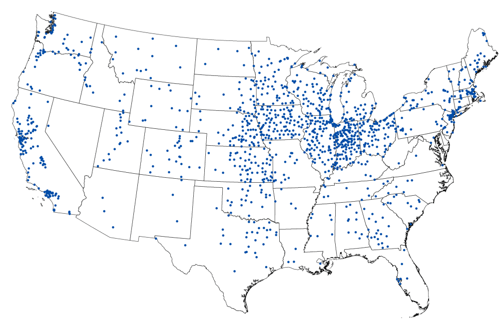
(a) Granted 1900 and before



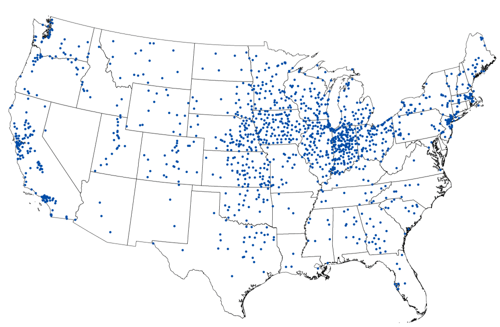
(b) Granted 1905 and before



(c) Granted 1910 and before



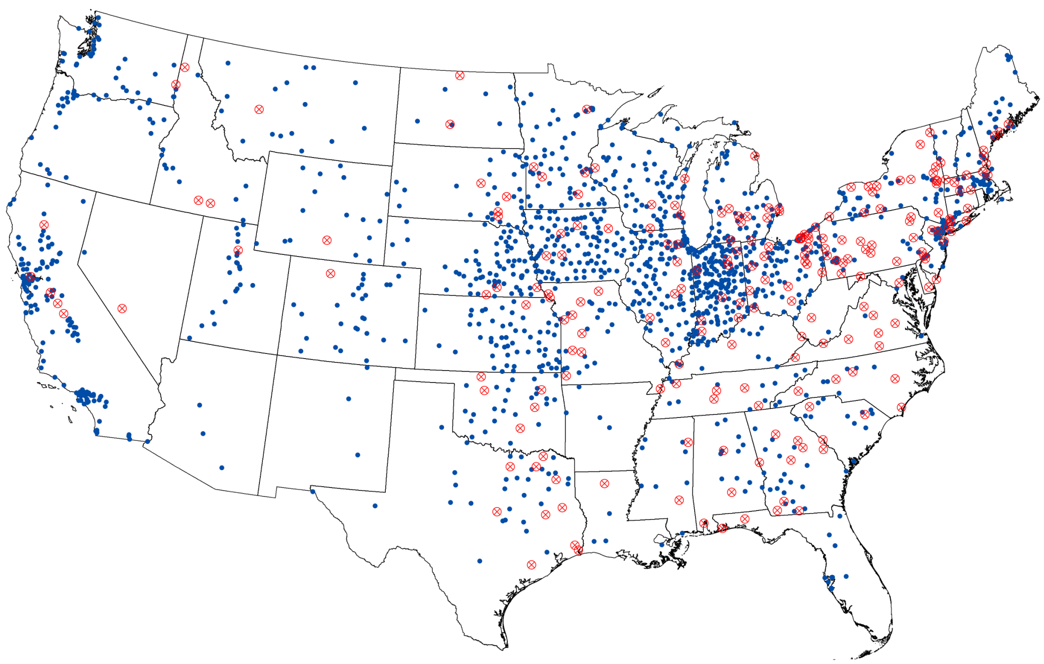
(d) Granted 1915 and before



(e) Granted 1920 and before (all libraries)

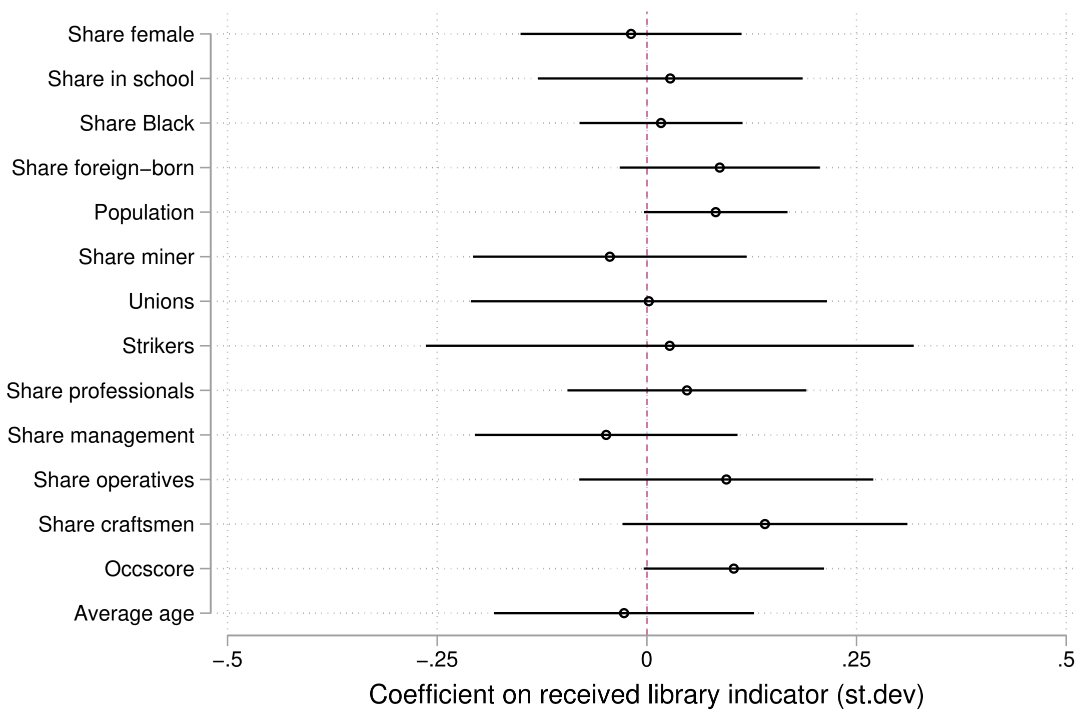
Notes: This figure shows the location of all cities that received a Carnegie library. Each panel shows the location of Carnegie libraries that were granted before the indicated dates. Source: Bobinski (1969) and Jones (1997).

Figure 5: **Map of all built and rejected Carnegie libraries**



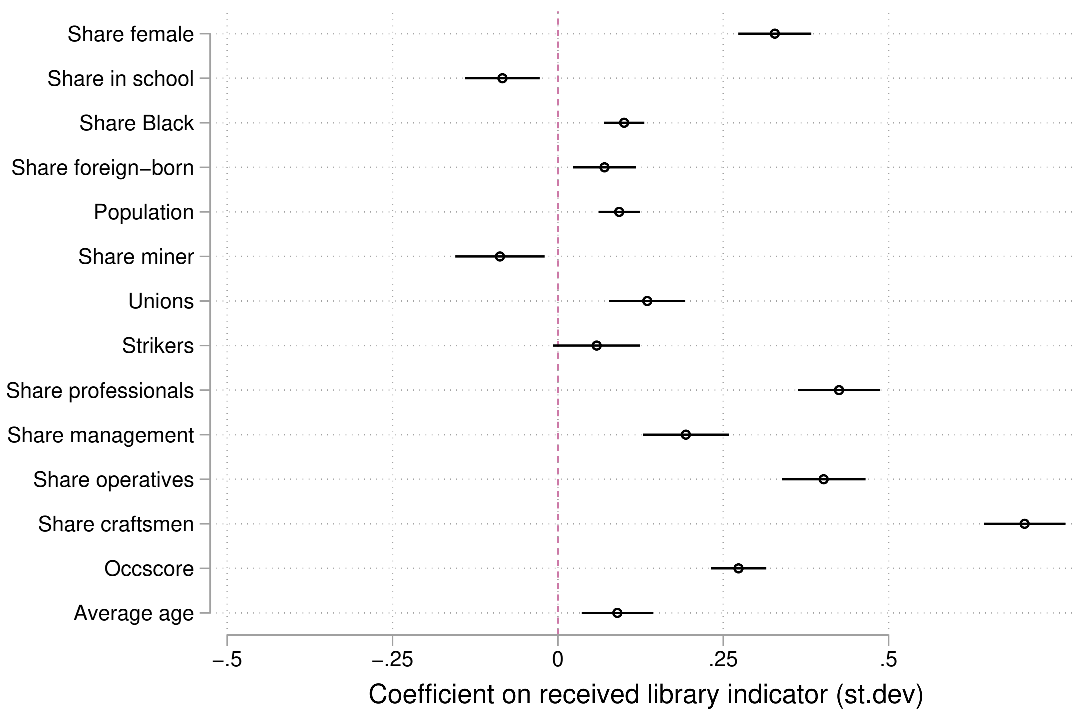
Notes: This figure shows the locations of all built and rejected Carnegie libraries. Darker dots correspond to libraries that were granted and built, as indicated in the previous figures. Red dots with an "X" correspond to cities that rejected their Carnegie library grant. Source: Bobinski (1969) and Jones (1997).

Figure 6: **Comparison of cities that accepted and rejected libraries**



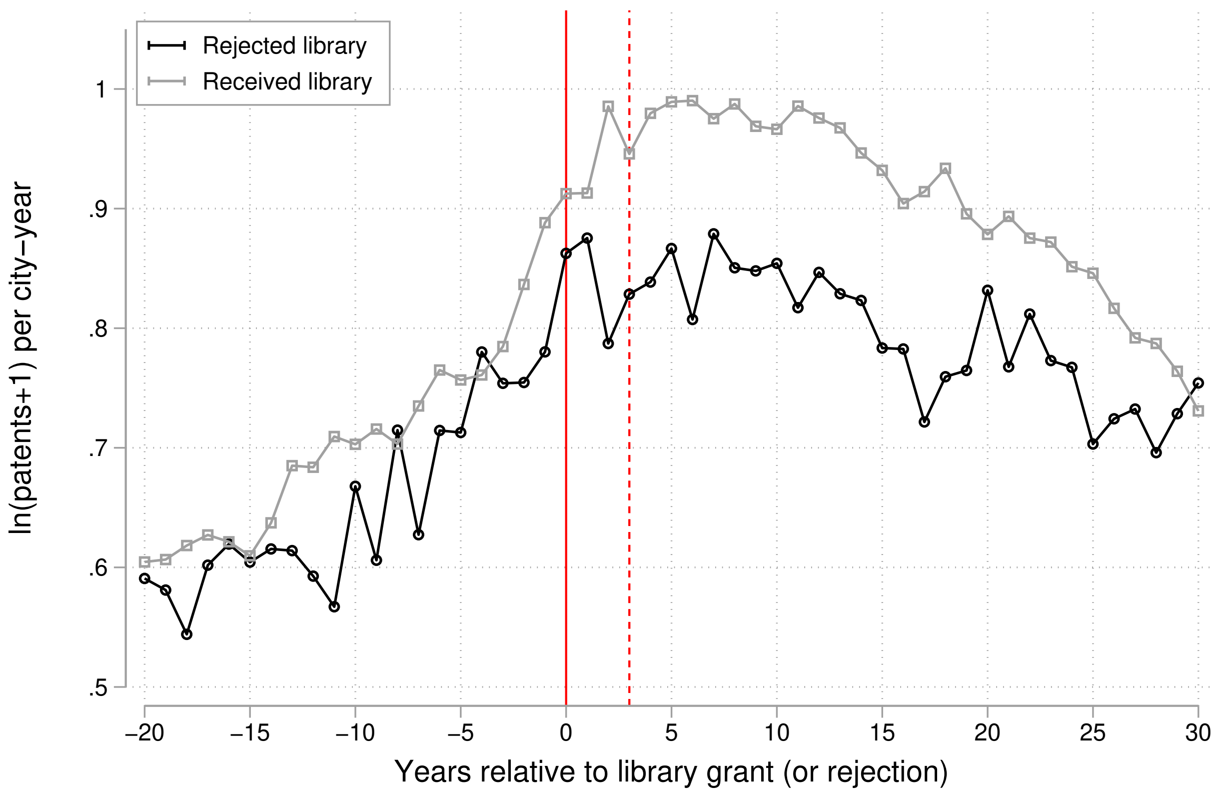
Notes: This figure shows comparisons between cities that accepted and rejected Carnegie libraries. We plot the standardized coefficient from a regression of each indicated city-level characteristic on a dummy variable for cities that built a Carnegie library conditional on state fixed effects. We standardize the covariates to have mean zero, standard deviation one and plot 95 percent confidence intervals. Covariates include population (measured in the library grant year after residualizing year fixed effects) and 1900 values for share of the population that is female, the share of the population in school, the share of the population that is Black, the share of the population in mining industries, occscore (a measure of average occupation-industry imputed earnings calculated using the Saavedra-Twinam algorithm), and the average age.

Figure 7: **Covariate comparisons across cities that accepted and did not apply for libraries**



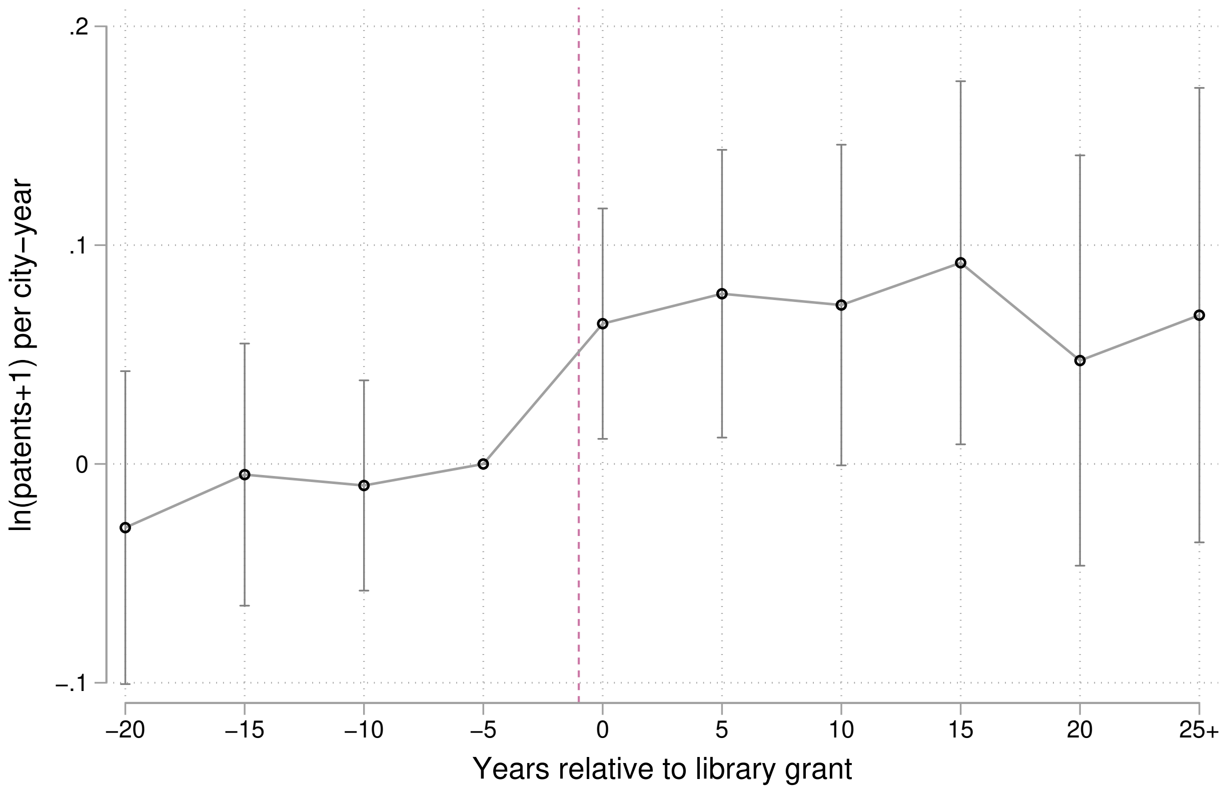
Notes: This figure shows comparisons between cities that accepted did not apply for Carnegie libraries. We plot the standardized coefficient from a regression of each indicated city-level characteristic on a dummy variable for cities that built a Carnegie library conditional on state fixed effects. We standardize the covariates to have mean zero, standard deviation one and plot 95 percent confidence intervals. Covariates include population (measured in the library grant year after residualizing year fixed effects) and 1900 values for the share of the population that is female, the share of the population in school, the share of the population that is Black, the share of the population in mining industries, occscore (a measure of average occupation-industry imputed earnings calculated using the Saavedra-Twinam algorithm), and the average age.

Figure 8: **Patents per city-year in cities that did and did not build libraries after library grants**



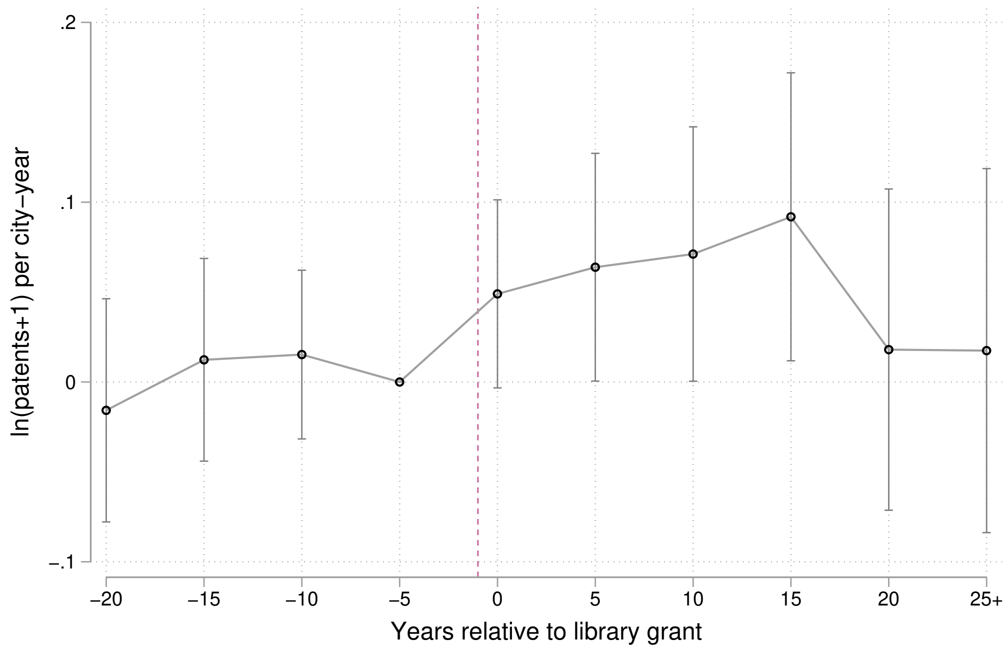
Notes: This figure shows the mean patenting counts across cities that received a Carnegie library and cities that were approved for a Carnegie grant but did not build a library after residualizing on grant date fixed effects. Averages are plotted by years relative to grant dates. The first line reflects library grant dates. The second dashed line illustrates that the mean time from library grants to when libraries were finished and opened to the public was three years.

Figure 9: **Event study estimates of Carnegie library receipt on city patenting**

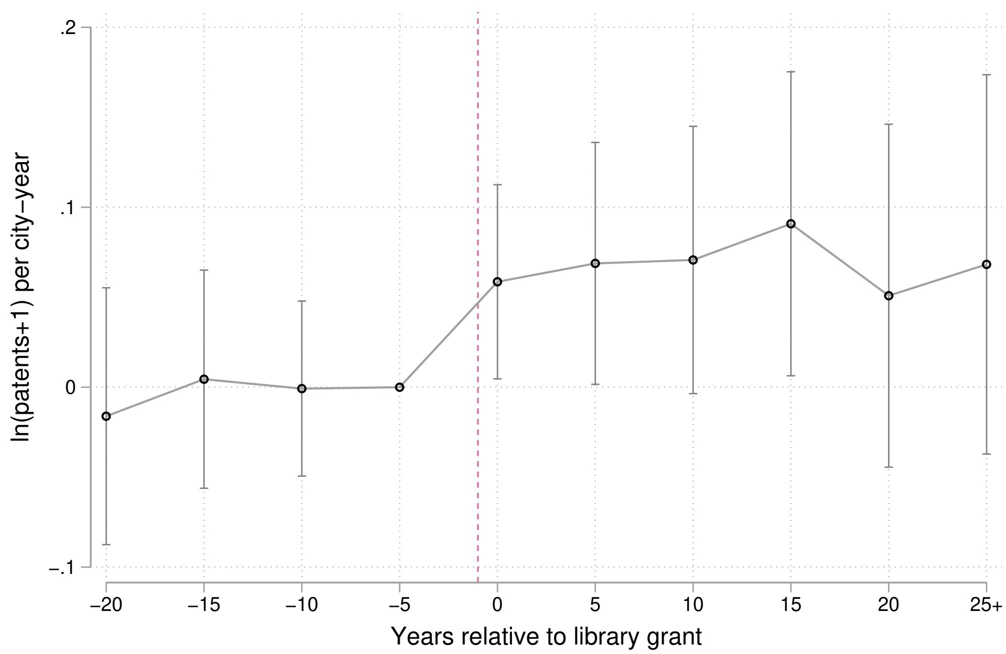


Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The reported coefficients reflect five year bins containing the labelled relative year and the four following years. For example, the bin labeled 5 contains the 5th-9th years after library grants. The coefficients are generated from interactions between a built library indicator and each five year increment, conditional on state-year and city fixed effects. The excluded category is 5 to 1 years before library grant dates. Standard errors are clustered at the city level and 90 percent confidence intervals are shown.

Figure 10: **Event study estimates of Carnegie library receipt on city patenting, alternative specifications**



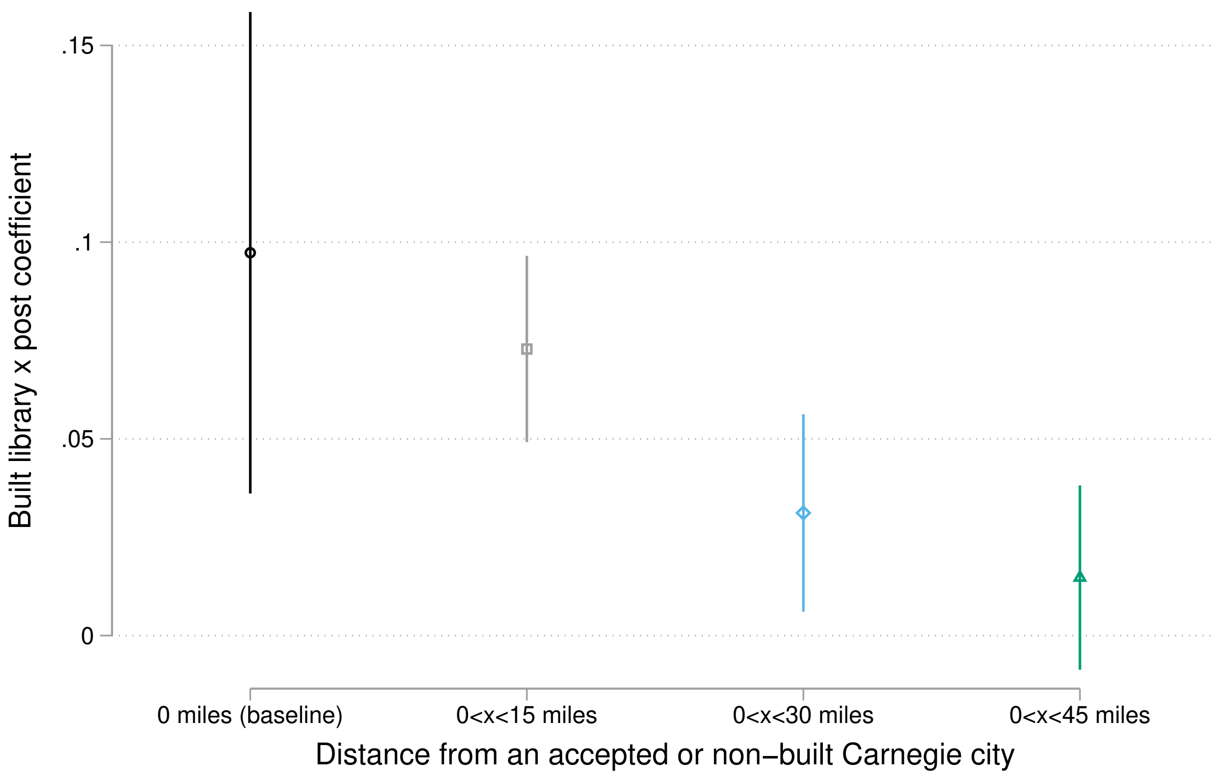
(a) Conditional on city and grant year-year fixed effects



(b) Conditional on city, grant year-year, and state-year fixed effects

Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The coefficients are generated from interactions between a build library indicator and each five year increment, conditional on the indicated fixed effects. Standard errors are clustered at the city level and 90 percent confidence intervals are graphed.

Figure 11: **Spillover effects of Carnegie libraries on patenting in nearby towns**



Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby town. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on city and state-year fixed effects. This estimate reflects a comparison of Carnegie and rejecting cities before and after grant dates. The remaining three estimates are for identically specified models with increasingly geographically dispersed treatment and control groups. In particular, treatment cities for the spillover regressions are defined as cities within the indicated number of miles from a Carnegie library. Control cities are defined as cities within the indicated number of miles from a rejected library and *not* within that distance from a Carnegie library. The Carnegie library receiving and rejecting cities themselves are excluded from all spillover samples. All standard errors are clustered at the grant-receiving-city level. 90 percent confidence intervals are shown.

Table 1: Effect of Carnegie libraries on patenting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.081	0.082	0.076	0.075	0.074	0.113	0.097	0.076
	(0.036)	(0.036)	(0.035)	(0.035)	(0.036)	(0.039)	(0.037)	(0.038)
Pre-1929 sample								
Built library × post	0.111	0.106	0.100	0.095	0.078	0.143	0.104	0.082
	(0.037)	(0.037)	(0.036)	(0.036)	(0.035)	(0.041)	(0.037)	(0.038)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean ln(pat+1)	0.817	0.817	0.817	0.817	0.817	0.817	0.817	0.817
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1, using a sample window of 20 years before and after Carnegie grants (first panel) or 20 years before Carnegie grants until 1929 (second panel). Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is $ln(patents + 1)$. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table 2: Effect of Carnegie libraries on measures of patent quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Had forward citation								
Built library × post	0.040	0.040	0.044	0.043	0.043	0.066	0.059	0.043
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
Count citations								
Built library × post	1.183	1.184	1.255	1.243	1.230	2.558	2.578	1.542
	(0.619)	(0.620)	(0.624)	(0.624)	(0.626)	(0.711)	(0.765)	(0.748)
Had top innovative pat.								
Built library × post	-0.009	-0.009	-0.008	-0.008	-0.008	0.010	0.007	-0.002
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean had forward citation	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.304
Mean forward citations	2.830	2.830	2.830	2.830	2.830	2.830	2.830	2.830
Mean had top innovative pat.	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.084
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on measures of patent quality. The estimates are from versions of Equation 1, using a sample window of 20 years before and after Carnegie grants. Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is the indicated measure of patent quality: either the probability of observing a patent that received a forward citation in the future, the count of forward citations that patents in that city-year generated, or the probability of observing a patent in the top 10 percentile of the measure developed by Kelly et al. (2021). Standard errors are shown in parentheses and clustered by city.

Table 3: Difference-in-differences estimates by patent classes

Cooperative Patent Classification	Library × post coefficient	Standard error
Human necessities	0.051	(0.018)
Performing ops/transport	0.093	(0.022)
Chemistry	0.013	(0.020)
Textiles	0.010	(0.010)
Constructions	0.043	(0.014)
Mech. engineering	0.075	(0.019)
Physics	0.017	(0.014)
Electricity	0.038	(0.013)

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. Each row represents a separate estimate of the baseline model in Equation 1 with $ln(patents + 1)$ for the indicated Cooperative Patent Classification as the outcome variable. All models include city and state-year fixed effects. Standard errors are clustered by city.

Table 4: Effect of Carnegie libraries on patents that cite past books, magazines, or patents

Dependent variable and sample	Library × post	Std. error	Mean of outcome
Observe a work-citing patent			
Full sample	0.0053	(0.0038)	0.0121
Pre-1929 sample	0.0061	(0.0037)	0.0117
Pre-1925 sample	0.0073	(0.0036)	0.0114
# work-citing patents			
Full sample	0.0110	(0.0066)	0.0171
Pre-1929 sample	0.0114	(0.0060)	0.0159
Pre-1925 sample	0.0119	(0.0059)	0.0154

Notes: This table shows results from our baseline model estimating the effect of Carnegie libraries on the probability of observing a prior-work citing patent and the number of patents that cite prior work. The procedure for identifying such patents is discussed in Section 3. All models include city and state-year fixed effects and standard errors are clustered by city.

Table 5: Effect of Carnegie libraries on multi-authored patenting

Dependent variable and sample	Library × post	Std. error	Mean of outcome
Observe a multi-author patent			
Full sample	0.0194	(0.0143)	0.1735
Pre-1929 sample	0.0221	(0.0147)	0.1760
Pre-1925 sample	0.0257	(0.0150)	0.1780
# multi-author patents			
Full sample	0.1132	(0.0401)	0.2637
Pre-1929 sample	0.1085	(0.0365)	0.2625
Pre-1925 sample	0.1143	(0.0369)	0.2640

Notes: This table shows results from our baseline model estimating the effect of Carnegie libraries on the probability of observing a multi-author patent and the number of multi-author patents. The procedure for identifying such patents is discussed in in Section 3. All models include city and state-year fixed effects and standard errors are clustered by city.

Table 6: Effect of Carnegie libraries on patenting, extensive margin ($Pat > 0$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library $\times$ post	0.025	0.025	0.022	0.022	0.022	0.015	0.005	0.014
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
Pre-1929 sample								
Built library $\times$ post	0.036	0.034	0.031	0.029	0.022	0.024	0.004	0.014
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)	(0.020)	(0.020)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean $Pr(Pat > 0)$	0.600	0.600	0.600	0.600	0.600	0.600	0.600	0.600
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1. Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is $Pr(Patents > 0)$. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table 7: Effect of Carnegie libraries on patenting, extensive margin (first-time inventors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First-time inventors								
Built library × post	0.294	0.294	0.287	0.283	0.280	0.379	0.356	0.205
	(0.102)	(0.102)	(0.101)	(0.101)	(0.102)	(0.104)	(0.105)	(0.109)
Share first-time								
Built library × post	-0.006	-0.008	-0.005	-0.005	-0.009	-0.011	-0.012	-0.008
	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)	(0.018)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean first-time count	1.350	1.350	1.350	1.350	1.350	1.350	1.350	1.350
Mean share first-time	0.626	0.626	0.626	0.626	0.626	0.626	0.626	0.626
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1. Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is indicated in each panel. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table 8: Effect of Carnegie libraries on women and immigrant patenting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women								
Built library × post	0.026	0.026	0.025	0.025	0.024	0.041	0.042	0.026
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.015)	(0.014)
Immigrants								
Built library × post	0.092	0.092	0.090	0.089	0.089	0.141	0.137	0.095
	(0.023)	(0.023)	(0.022)	(0.022)	(0.022)	(0.028)	(0.031)	(0.030)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean women patents	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
Mean immigrant patents	0.226	0.226	0.226	0.226	0.226	0.226	0.226	0.226
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1, using a sample window of 20 years before and after Carnegie grants. Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is the count of patents for women (first panel) and immigrants (second panel), each identified using the name-based procedure described in Section 3. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table 9: Heterogeneity in library difference-in-differences estimates across city characteristics

Coefficient	Estimate	Standard error
Built library × post × had college	0.039	(0.141)
Built library × post × share in-school top half	0.084	(0.067)
Built library × post × imputed income top half	0.017	(0.070)
Built library × post × share craftsmen top half	0.071	(0.073)
Built library × post × share Black top half	0.008	(0.068)
Built library × post × population top half	0.114	(0.073)

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. Each row represents a separate estimate of the baseline model in Equation 1 with $ln(patents + 1)$ for the as the outcome variable. All models include city and state-year fixed effects. Coefficients are the triple interaction between each indicated variable, an indicator for whether the city built a Carnegie library, and an indicator variable for years after library grants were made. All interactions are fully saturated—each model also includes the indicated city characteristic, the interaction of the indicated city characteristic and the post variable, and the interaction of the city characteristic with the built library variable. Standard errors are clustered by city.

Table 10: Robustness of difference-in-differences patent results to alternative samples

Specification	Library × post	Std. error	Cities
Baseline model	0.097	(0.037)	1,368
Excluding observations from the baseline sample			
Exclude Southern states	0.120	(0.041)	1,196
Exclude New York and Pennsylvania	0.082	(0.038)	1,289
15 year pre-period	0.086	(0.036)	1,368
10 year pre-period	0.085	(0.036)	1,368
Exclude "finance" motivated rejectors	0.093	(0.045)	1,318
Exclude cities larger than 15,000 people	0.080	(0.038)	1,267
Exclude cities larger than 5,000 people	0.079	(0.039)	922
Adding observations to the baseline sample			
Include pre-1899 grants	0.092	(0.037)	1,374
Include large population cities/counties	0.120	(0.038)	1,466
Include control cities with local library philanthropists	0.092	(0.035)	1,389
Include cities missing 1900 covariates	0.072	(0.038)	1,470
Include all cities	0.089	(0.036)	1,602

*Notes:*This table shows results from our baseline model with state-year and city fixed effects estimated on different samples of the data. The first panel shows results after excluding observations from the baseline sample, while the second panel shows results after including additional observations into the baseline sample. The 15 and 10-year pre-period labels indicate regressions where we restrict the pre-period to 15 and 10 years before the libraries were granted, instead of the 20 year pre-period that we use in our baseline analysis.

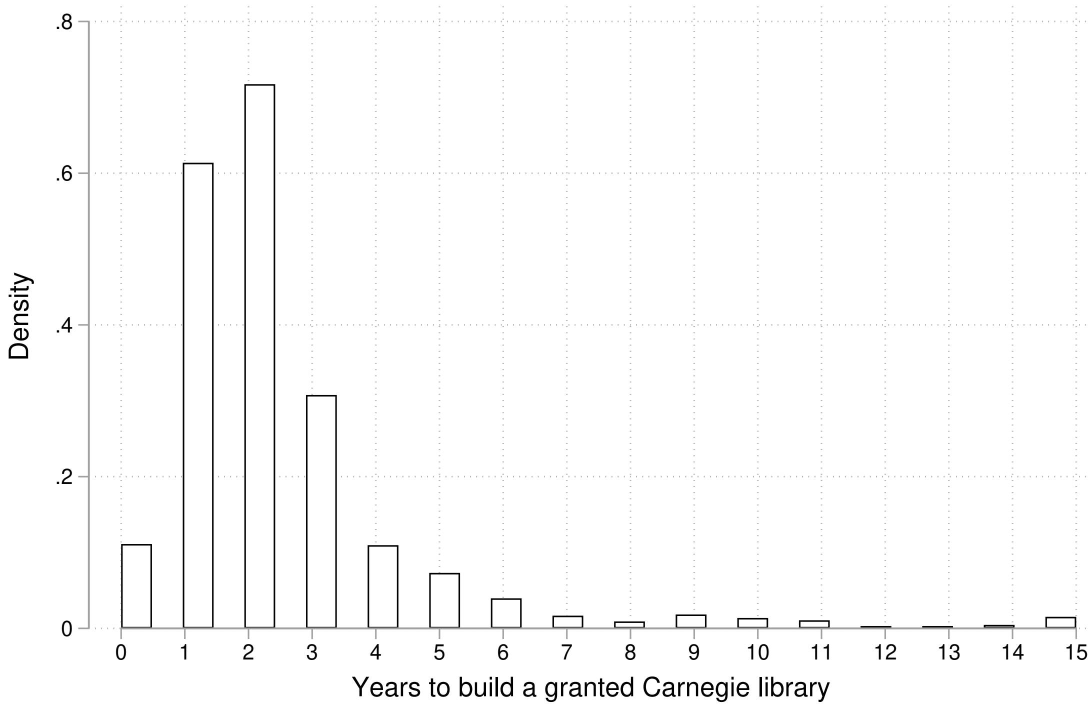
Table 11: Robustness of difference-in-differences results to alternative patent measures and estimation strategies

Dependent variable and estimation strategy	Library $\times$ post	Std. error	Cities
Specification robustness			
Baseline $ln(pat+1)$ model	0.097	(0.037)	1,368
Condition on time-varying log population	0.087	(0.035)	1,353
Condition on Post $\times$ Strikers	0.094	(0.037)	1,368
Condition on Post $\times$ KoL Assembles	0.094	(0.037)	1,368
Condition on Post $\times$ Strikers & Post $\times$ KoL Assembles	0.094	(0.038)	1,368
Condition on Post $\times$ Had College	0.092	(0.037)	1,368
Condition on Post $\times$ Share kids in school (1900)	0.100	(0.037)	1,368
Outcome measurement robustness			
Patent measure: $ihs(patents)$	0.112	(0.046)	1,368
Patent measure: patent counts	1.375	(0.334)	1,368
Marginal effect (count) from Poisson model	0.530	(0.291)	1,368
Marginal effect (count) from Poisson model, no city FE	1.257	(0.255)	1,368
Marginal effect (count) from Zero-inflated Poisson model	0.897	(0.252)	1,368
Marginal effect (count) from Negative Binomial model	0.621	(0.233)	1,368

Notes: This table shows results from our baseline model using different outcome variables and estimation strategies. KoL stands for the Knights of Labor. The number of pre-1900 strikers and KoL assemblies are measured at the county level in logs and are measured using the data collected by Bittarello (2019). The variable Had College is an indicator for whether a city had a college in 1900, sourced from the 1900-1901 academic year version of the Report of the Commission of Education published by the U.S. Bureau of Education. The baseline model conditions on state-year and city fixed effects. The zero-inflated Poisson and negative binomial models exclude state-year fixed effects to allow convergence. All standard errors are clustered by city.

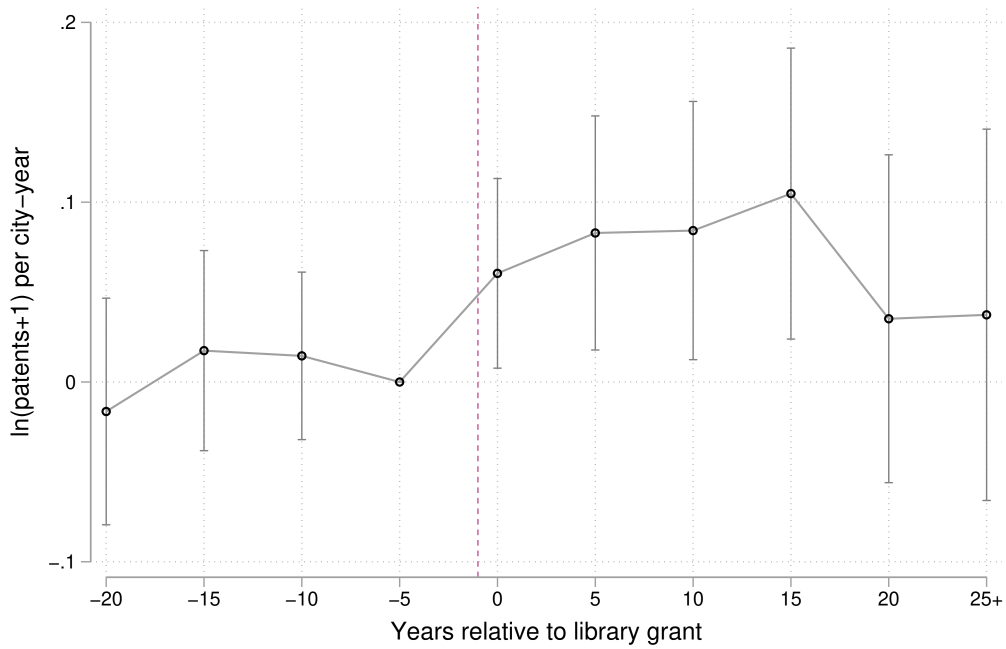
A Figures and tables for online publication

Figure A1: **Distribution of time required to construct libraries after library grants**

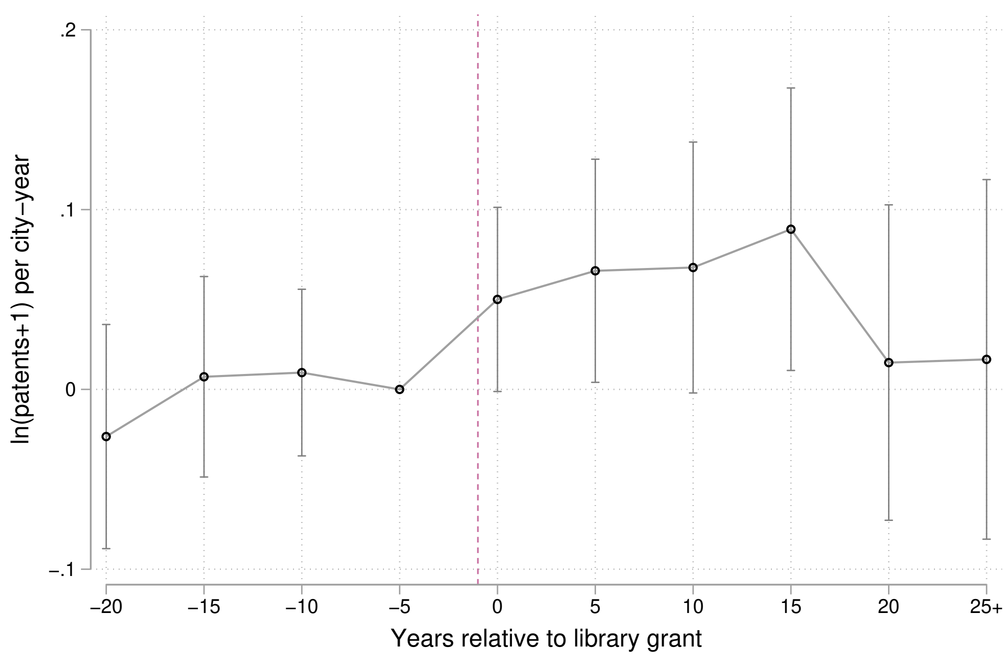


Notes: This figure shows the distribution of time (measured in years) between library grants and the libraries being opened to the public. The distribution is top-coded at 15 years. The average construction time was 3 years but the most frequent construction times were 1 or 2 years.

Figure A2: **Event study estimates of Carnegie library receipt on city patenting, alternative log specifications**



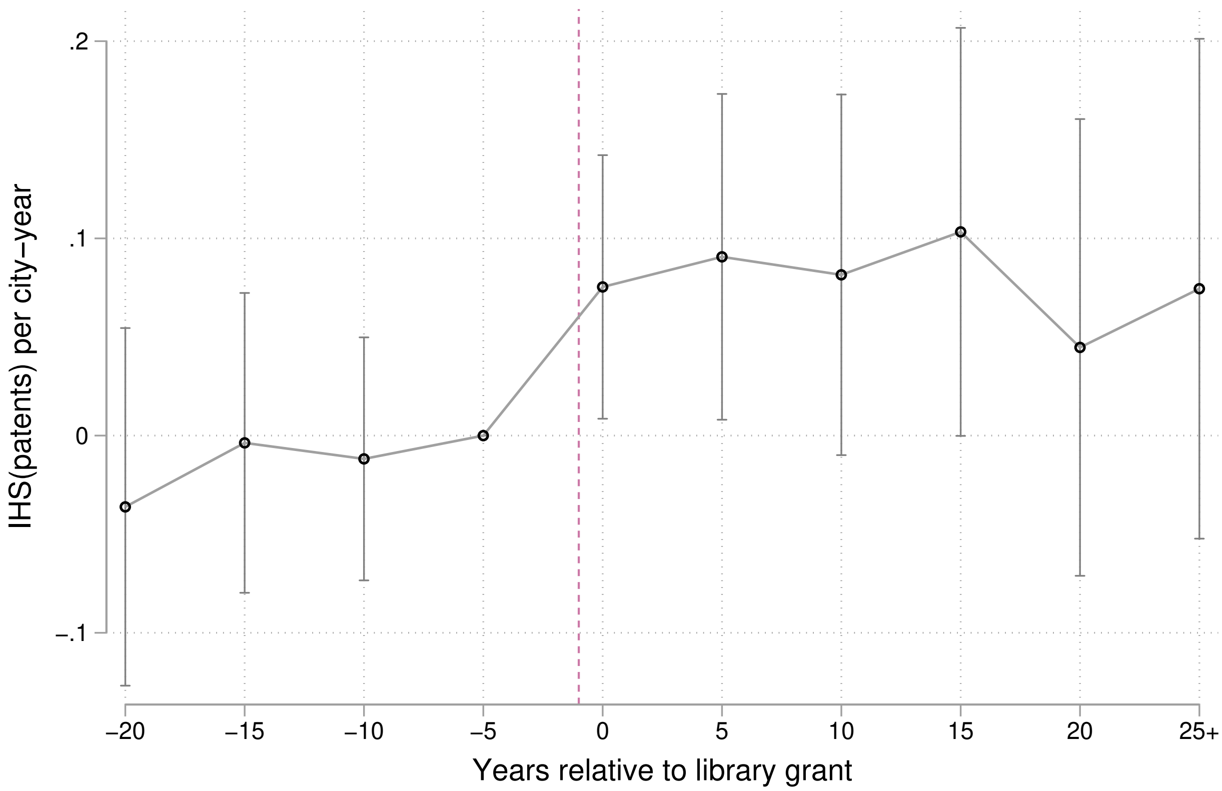
(a) Conditional on city fixed effects



(b) Conditional on city and year fixed effects

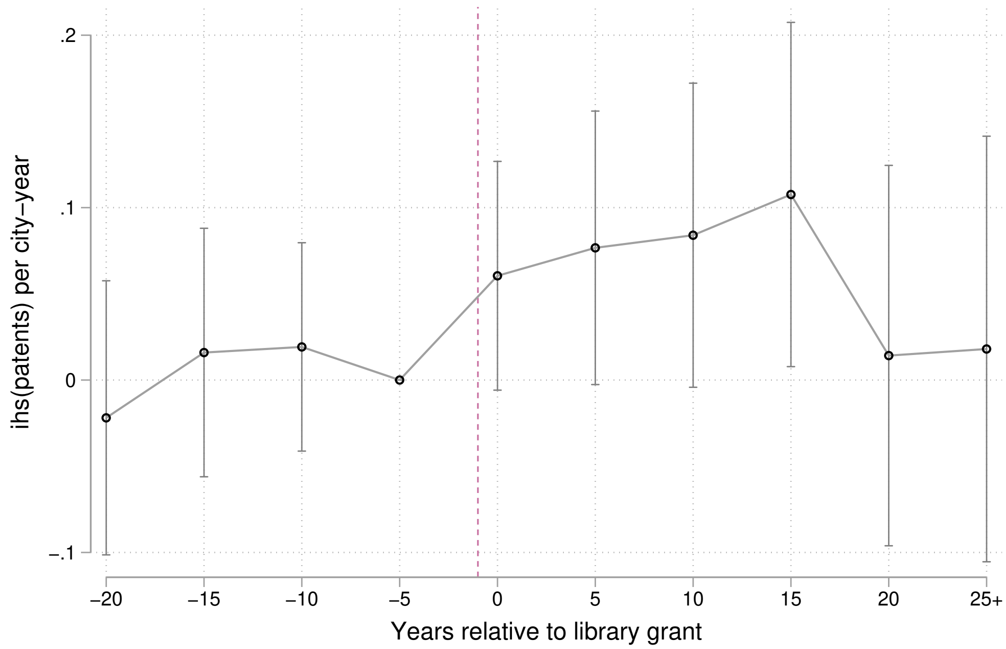
Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The coefficients are generated from interactions between a built library indicator and each five year increment, conditional on the indicated fixed effects.

Figure A3: **Event study estimates of Carnegie library receipt on city patenting, inverse hyperbolic sine specification**

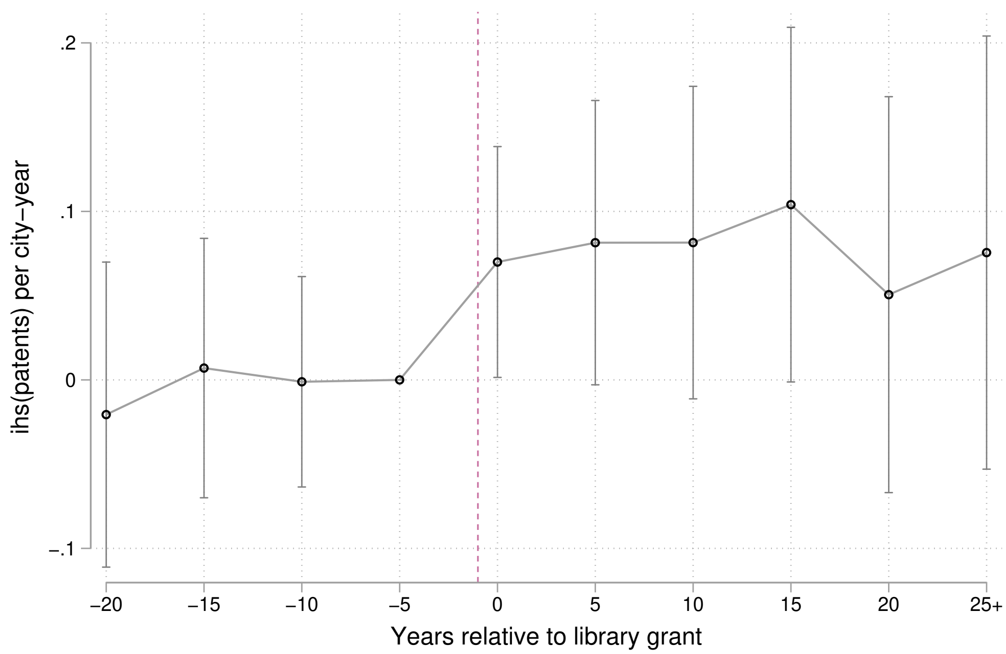


Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The reported coefficients reflect five year bins containing the labelled relative year and the four following years. For example, the bin labeled 5 contains the 5th-9th years after library grants. The coefficients are generated from interactions between a built library indicator and each five year increment, conditional on state-year and city fixed effects. The excluded category is 5 to 1 years before library grant dates. Standard errors are clustered by cities and 90 percent confidence intervals are shown.

Figure A4: **Event study estimates of Carnegie library receipt on city patenting, alternative inverse hyperbolic sine specifications**



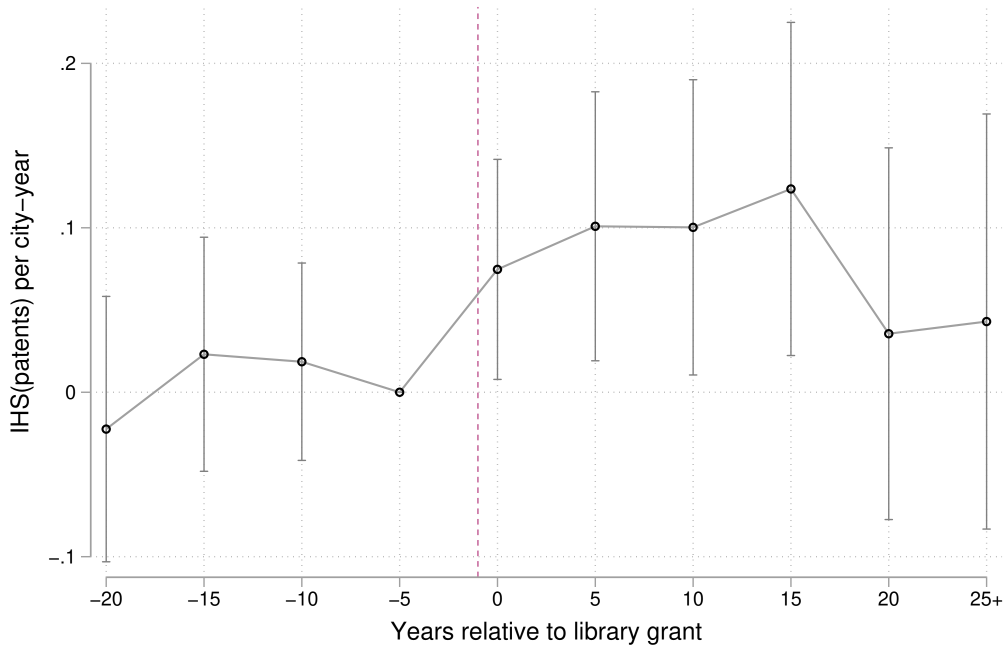
(a) Conditional on city and grant year-year fixed effects



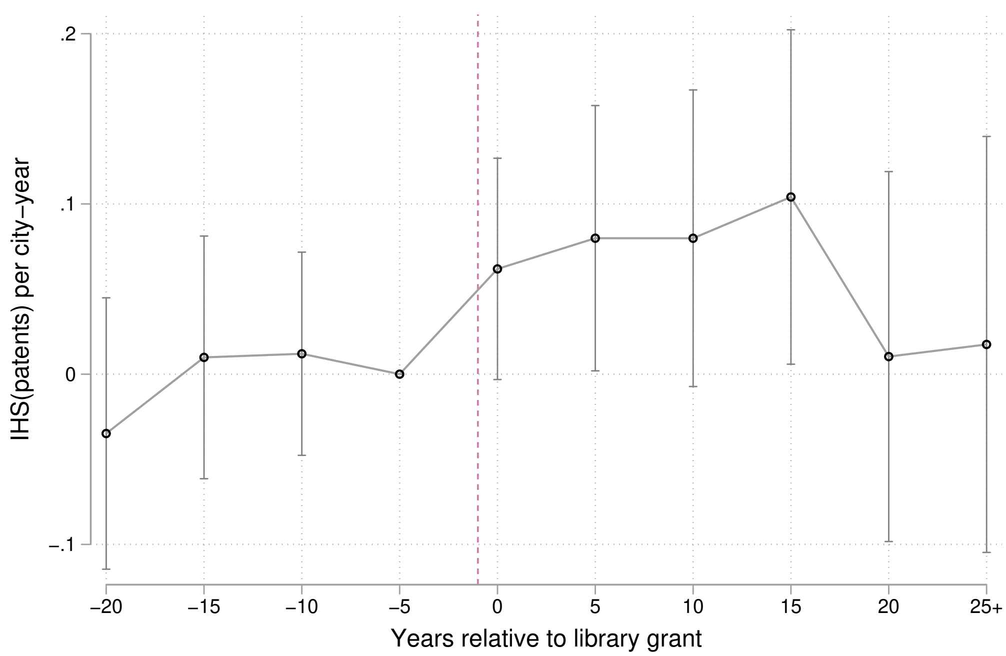
(b) Conditional on city, grant year-year, and state-year fixed effects

Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The coefficients are generated from interactions between a built library indicator and each five year increment, conditional on the indicated fixed effects. Standard errors are clustered at the city level and 90 percent confidence intervals are shown

Figure A5: **Event study estimates of Carnegie library receipt on city patenting, additional inverse hyperbolic sine specifications**



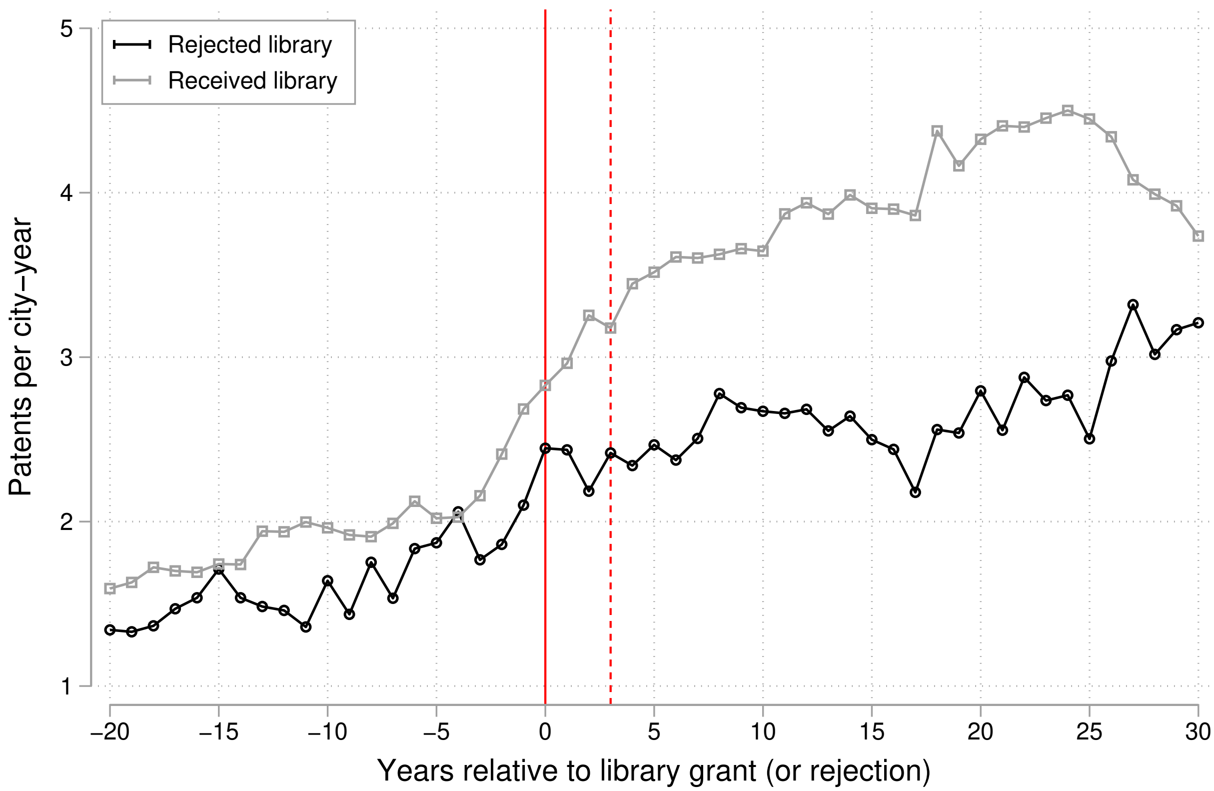
(a) Conditional on city fixed effects



(b) Conditional on city and year fixed effects

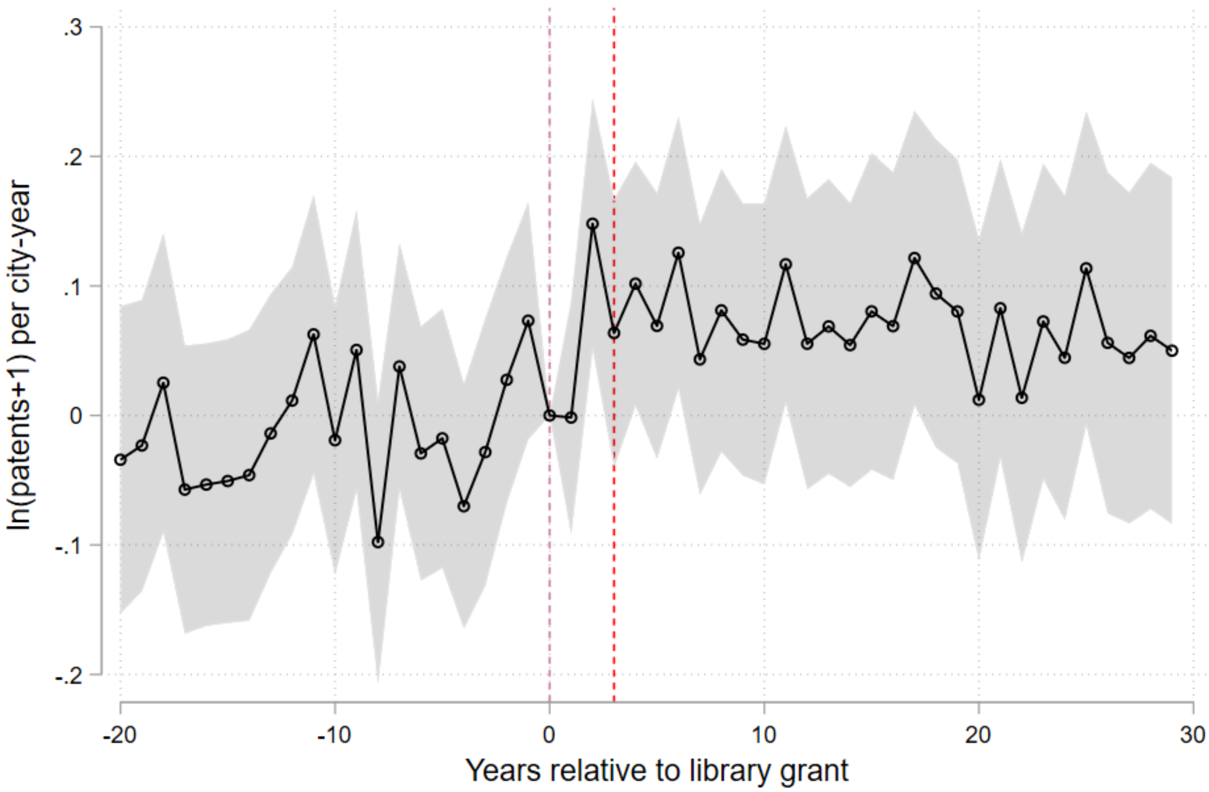
Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The coefficients are generated from interactions between a built library indicator and each five year increment, conditional on the indicated fixed effects. Standard errors are clustered at the city level and 90 percent confidence intervals are shown

Figure A6: **Patents per city-year in cities that did and did not build libraries after library grants, patent counts**



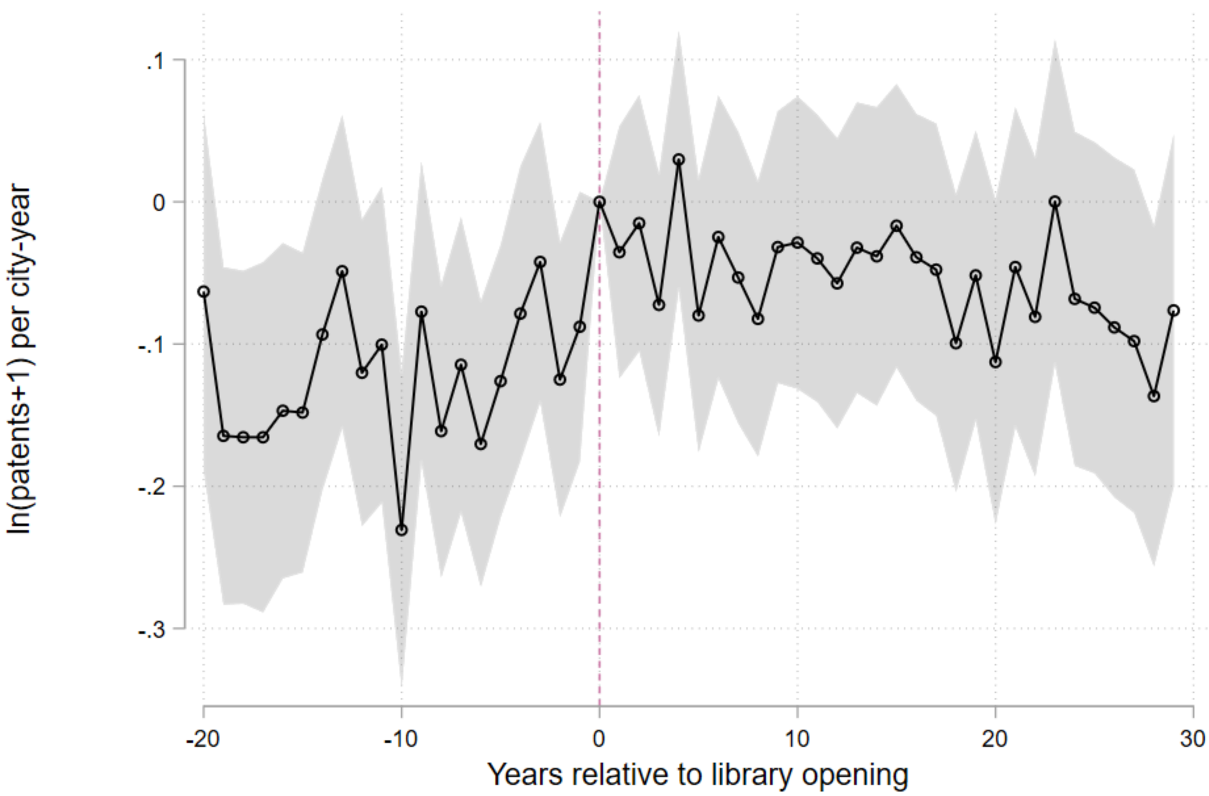
Notes: This figure shows the mean patenting counts across cities that received a Carnegie library and cities that were approved for a Carnegie grant but did not build a library after residualizing on grant date fixed effects. Averages are plotted by years relative to grant dates. The first line reflects library grant dates. The second dashed line illustrates that the mean time from library grants to when libraries were finished and opened to the public was three years.

Figure A7: **Event study estimates of Carnegie library receipt on city patenting, yearly estimates**



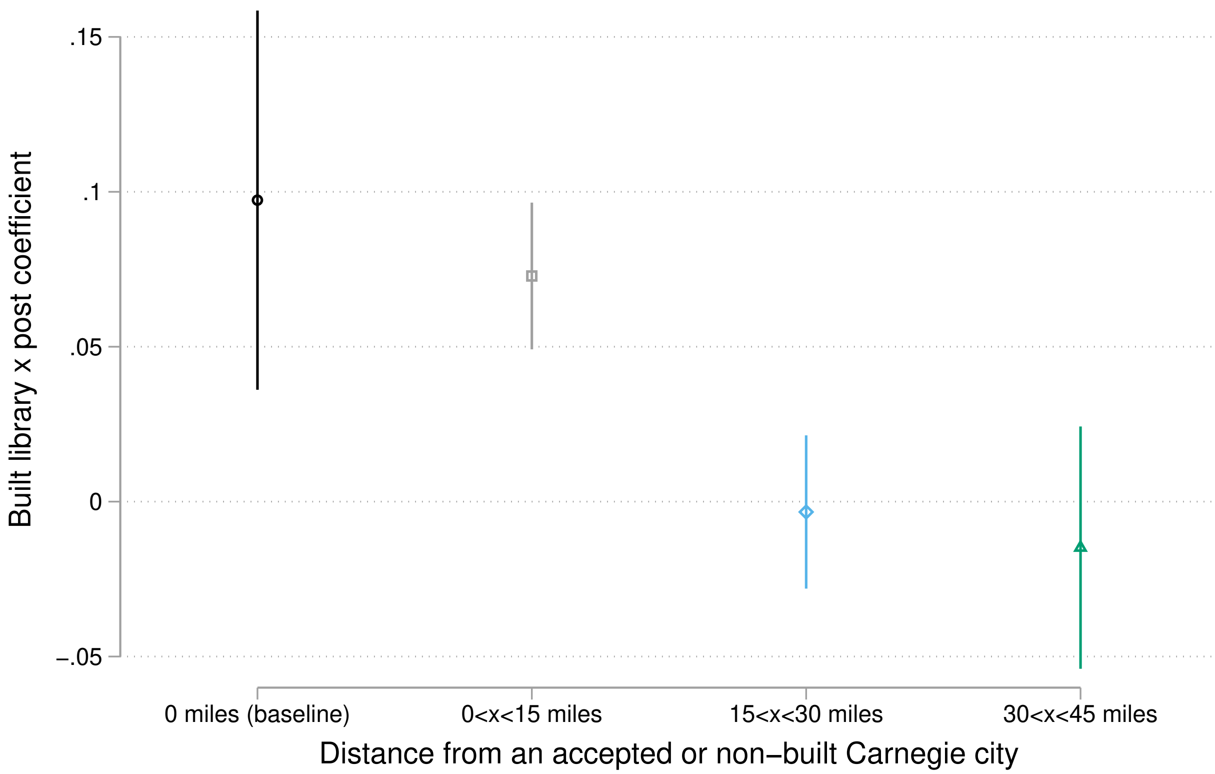
Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library grant dates. The coefficients are generated from interactions between a built library indicator and relative year dummy variables, conditional on state-year and city fixed effects. The excluded category is the year of the library grants. The red dashed line indicates that the mean time from library grants to when libraries were finished and opened to the public was three years. Standard errors are clustered at the city level and 90 percent confidence intervals are shown.

Figure A8: **Event study estimates of Carnegie library receipt on city patenting, yearly estimates using opening dates**



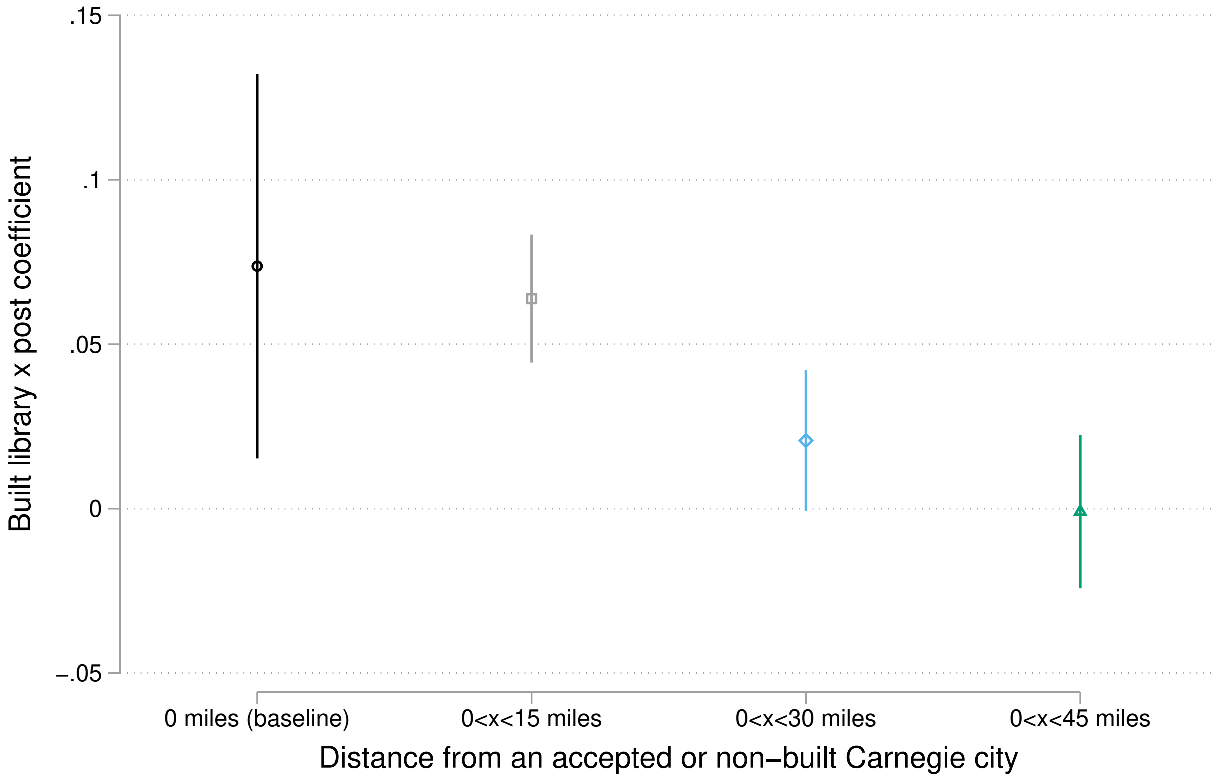
Notes: This figure shows the marginal effects of receiving a Carnegie library relative to rejecting a library on patenting in the years before and after library opening dates. For rejecting cities and libraries missing build days, we impute the build time as two years (the median of non-missing observations). The reported coefficients reflect five year bins containing the labelled relative year. The coefficients are generated from interactions between a built library indicator and relative year dummy variables, conditional on state-year and city fixed effects. The excluded category is the year of the library opening. Standard errors are clustered at the city level and 90 percent confidence intervals are shown.

Figure A9: **Spillover effects of Carnegie libraries on patenting in nearby towns, non-intersecting distance bins**



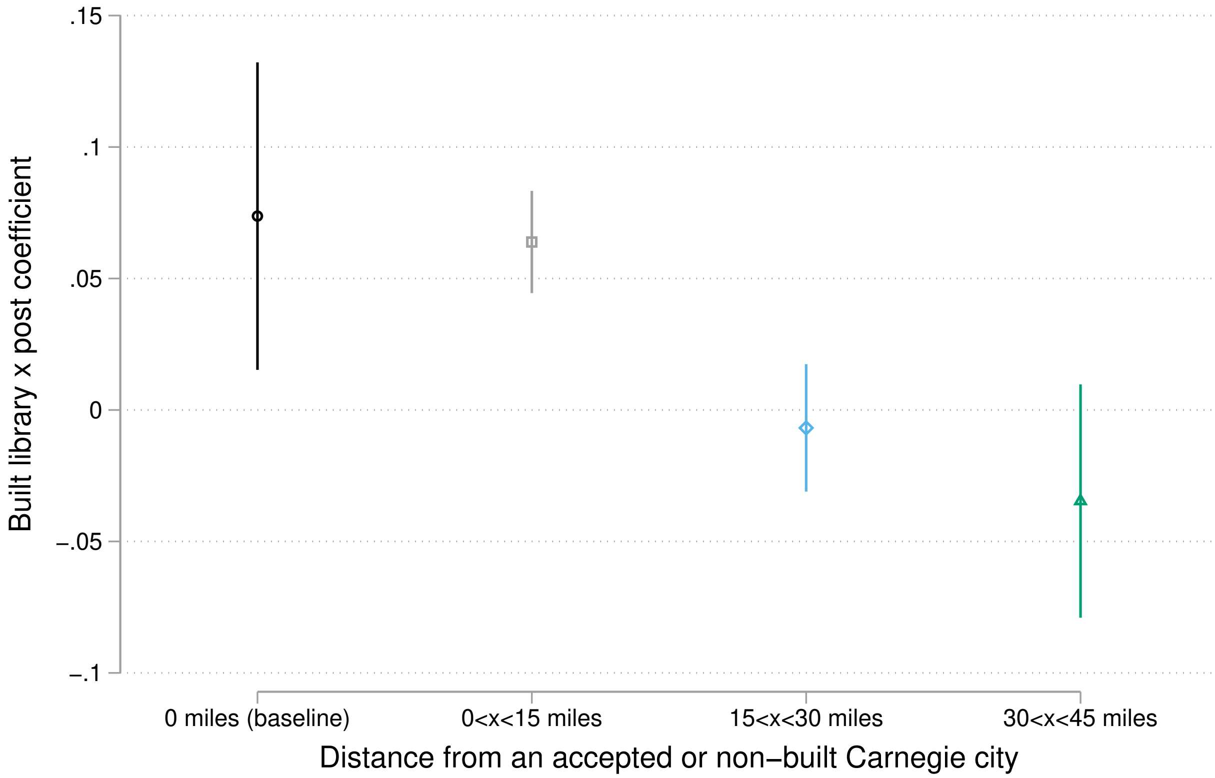
Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby town. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on city and state-year fixed effects. This estimate reflects a comparison of Carnegie and rejecting cities before and after grant dates. The remaining three estimates are for identically specified models with increasingly geographically dispersed treatment and control groups. In particular, treatment cities for the spillover regressions are defined as cities within the indicated number of miles from a Carnegie library. Control cities are defined as cities within the indicated number of miles from a rejected library and *not* within that distance from a Carnegie library. The Carnegie library receiving and rejecting cities themselves are excluded from all spillover samples. All standard errors are clustered at the grant-receiving-city level. 90 percent confidence intervals are shown.

Figure A10: **Spillover effects of Carnegie libraries on patenting in nearby towns, city and year fixed effects**



Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby town. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on city and year fixed effects. This estimate reflects a comparison of Carnegie and rejecting cities before and after grant dates. The remaining three estimates are for identically specified models with increasingly geographically dispersed treatment and control groups. In particular, treatment cities for the spillover regressions are defined as cities within the indicated number of miles from a Carnegie library. Control cities are defined as cities within the indicated number of miles from a rejected library and *not* within that distance from a Carnegie library. The Carnegie library receiving and rejecting cities themselves are excluded from all spillover samples. All standard errors are clustered at the grant-receiving-city level. 90 percent confidence intervals are shown.

Figure A11: **Spillover effects of Carnegie libraries on patenting in nearby towns, city and year fixed effects, non-intersecting distance bins**



Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby town. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on city and year fixed effects. This estimate reflects a comparison of Carnegie and rejecting cities before and after grant dates. The remaining three estimates are for identically specified models with increasingly geographically dispersed treatment and control groups. In particular, treatment cities for the spillover regressions are defined as cities within the indicated number of miles from a Carnegie library. Control cities are defined as cities within the indicated number of miles from a rejected library and *not* within that distance from a Carnegie library. The Carnegie library receiving and rejecting cities themselves are excluded from all spillover samples. All standard errors are clustered at the grant-receiving-city level. 90 percent confidence intervals are shown.

Table A1: Summary statistics

	Mean	Std. dev.	Min.	Max.
City-year variables (N = 54,573)				
$ln(patents + 1)$	0.817	0.870	0.000	5.207
$ihs(patents)$	1.042	1.087	0.000	5.894
Patent count	2.715	6.126	0.000	181.500
Female patents	0.100	0.340	0.000	14.125
Immigrant patents	0.226	0.565	0.000	16.959
Observed a patent that cited prior work	0.012	0.109	0.000	1.000
Patents that cite prior work	0.017	0.191	0.000	18.000
Observed a multi-inventor patent	0.173	0.379	0.000	1.000
Multi-inventor patents	0.264	0.872	0.000	37.000
Forward citations	2.830	12.381	0.000	694.833
Had a forward citation	0.304	0.460	0.000	1.000
Had a p90 breakthrough patent	0.084	0.278	0.000	1.000
Count first-time patents	1.350	2.475	0.000	61.500
1900 time-invariant variables (N = 1,368)				
Built Carnegie library	0.877	0.328	0.000	1.000
Share in-school	0.619	0.127	0.025	0.919
Share Black	0.055	0.135	0.000	1.000
Share female	0.493	0.036	0.204	0.570
Share miner	0.016	0.058	0.000	0.708
Occupation-industry earnings proxy	17.982	2.495	5.336	24.800
Population (log)	8.122	0.971	4.419	10.295
Average age	27.386	2.658	18.759	40.471
Share professional	0.079	0.033	0.000	0.316
Share managers	0.091	0.040	0.000	0.293
Share craftsmen	0.151	0.058	0.000	0.412
Share skilled operators	0.136	0.086	0.000	0.752
Strikers (county)	536.099	3401.175	0.000	76949.000
Knights of Labor Assemblies (county)	2.656	9.565	0.000	152.000
Had college	0.097	0.296	0.000	1.000

Notes: This table shows summary statistics for the main variables and sample used in this paper. The construction of these data is described in Section 3 and Appendix B.

Table A2: Effect of Carnegie libraries on patenting, inverse hyperbolic sine

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.097	0.098	0.091	0.089	0.088	0.133	0.112	0.089
	(0.045)	(0.045)	(0.044)	(0.044)	(0.044)	(0.048)	(0.046)	(0.047)
Pre-1929 sample								
Built library × post	0.135	0.128	0.120	0.115	0.094	0.169	0.120	0.096
	(0.046)	(0.046)	(0.045)	(0.044)	(0.044)	(0.050)	(0.046)	(0.047)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean ihs(pat)	1.042	1.042	1.042	1.042	1.042	1.042	1.042	1.042
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1, using a sample window of 20 years before and after Carnegie grants (first panel) or 20 years before Carnegie grants until 1929 (second panel). Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is $ihs(patents)$. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table A3: Effect of Carnegie libraries on patenting, patent counts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.846	0.848	0.825	0.814	0.805	1.379	1.375	0.970
	(0.255)	(0.255)	(0.251)	(0.251)	(0.252)	(0.315)	(0.335)	(0.328)
Pre-1929 sample								
Built library × post	0.988	0.958	0.934	0.901	0.822	1.525	1.409	0.987
	(0.272)	(0.270)	(0.267)	(0.266)	(0.258)	(0.336)	(0.343)	(0.336)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean patents	2.715	2.715	2.715	2.715	2.715	2.715	2.715	2.715
Observations	54,573	54,573	54,573	54,573	54,573	54,573	54,573	54,573
Cities	1,368	1,368	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates are from versions of Equation 1, using a sample window of 20 years before and after Carnegie grants (first panel) or 20 years before Carnegie grants until 1929 (second panel). Built library indicates cities that built a Carnegie library. Post indicates years after cities received Carnegie library grants. Post is defined for all cities, since every city in our sample received a Carnegie library grant. If a city received multiple grants, Post indicates the earliest grant year. The outcome variable is the count of patents. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table A4: Effect of Carnegie libraries on patenting (aggregated model)

	$ln(pat + 1)$		$ihs(pat)$		Poisson (count)	
Built library × post	0.131	0.185	0.143	0.198	11.996	14.548
	(0.072)	(0.075)	(0.080)	(0.084)	(5.598)	(6.004)
City FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
State FE - Post FE		✓		✓		✓
Mean patent measure	3.163	3.163	3.766	3.766	54.147	54.147
Observations	2,736	2,736	2,736	2,736	2,736	2,736
Cities	1,368	1,368	1,368	1,368	1,368	1,368

Notes: This table shows the impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. The estimates in this table are from a model that sums all patents in each city before and after they received a library grant, such that each city in the sample has two observations. The different columns correspond to alternative outcome variable transformations or estimation strategies that are used on the aggregated data.

B Data appendix

In this appendix we provide additional information on the data used in this project. This appendix supplements Section 3, which describes the core details of the data construction.

B.1 Library data

Our data on Carnegie library locations comes from multiple sources. The first two are books by library historians. First, we use Bobinski (1969), who was the first to systematically record the location of Carnegie libraries in his seminal work. We compare this list to our second source, Jones (1997)—who identified a few additional libraries and updated the grant dates in a handful of cases. When dates differ, we undertake research using library websites to determine who was correct, and use the proper date. Finally, as a check, we compare our data to other compiled sources of Carnegie libraries. These sources include Wikipedia—where editors maintain information about each Carnegie library in the U.S., including whether or not the building is still a library today — and regional websites like "California Libraries of California."[45] Since the primary source of these compilations are the same library historians that we use, our data aligns with those sources. Our data on rejected libraries—and the reasons cities may have not built libraries—comes from Bobinski (1969), supplemented by our research using the original Carnegie library correspondence. This correspondence was obtained from the Carnegie Collations at the Rare Book and Manuscript Library at Columbia University in New York City.

A limitation of prior scholarship on Carnegie libraries is dating library openings. While we have excellent data on when Carnegie approved cities for library grants—because letters were standardized and were archived by the Carnegie Corporation—data on when libraries were actually constructed and open to the public was not centrally tracked. This timing is important for our analysis, since we would not expect to see any effects of library construction on innovation before libraries actually opened. We compile new data on the universe of library openings by searching library websites, state historical associations, and newspapers. Since the last Carnegie grant was given in 1919, most still-standing buildings have celebrated their 100th anniversary in the last two decades. These anniversaries typically generate local newspaper stories which contain information about the

[45] The website is https://www.carnegie-libraries.org

opening date, and are a primary source for our opening dates. We also called a number of libraries and received information directly from staff. A few libraries (roughly 5 percent of the total) could not be assigned opening dates. These are most often small libraries that were torn down. We use information on library opening dates to calculate the average time to library opening (3 years) and confirm that the timing of our effects is consistent with construction timing.

B.2 Patent data

Inventor names disambiguation

We assign each inventor who filed a patent between 1860 and 1960 a unique ID based on name similarity and their city of residence as reported in the patent text. For each inventor, we check if there is another inventor whose first names start with the same letter and whose full names satisfy the following condition:

$$round\left(\left(2 * \frac{M}{T}\right) \times 100\right) \geq 90$$

where M is the number of matches and T is the length of both names. We only match names if they filed a patent within 10 years and lived within 50 km of each other or within 5 years and lived within 400 km of each other. We do this to take into consideration, particularly at the beginning of the sample, inventors who moved from the country to the city. The unique IDs allow us to identify the first year in which each inventor filed a patent. A limitation of this approach is that if inventors move further away, we risk re-classifying them as a new inventor. On the other hand, setting a larger distance radius risks combining otherwise distinct inventions with the name same. Manual inspections suggest that the 400km rule strikes a good balance between these two concerns, and we have verified that our central findings are not sensitive to alternative, reasonable choices of a distance threshold.

Identify citing patents

To identify the patents that cite prior materials (books, magazines, patents), we first identify a set of terms that are likely to be associated to those materials (e.g., encyclopedia, handbook, dictionary, etc.). We then search the corpus of patents and identify those that mention these keywords. We manually review the matches and extend the set of keywords based on our matches. This process is repeated until we are unable to identify new terms. We report below the list of keywords that we identified, together with the punctuation and matching rules that allow us to minimize false positives. Often we require that a keyword is followed by a number (e.g., page). When we do so, we match when the numbers are multiples of five. This is because historical patents report line numbers (in multiples of 5) in the margins, and the OCR process often digitizes and adds them to the text.

Our final keywords are: *pages* (or *pp.*) followed by a number; *volume* (*vol.*, *vol;*, or *yol.*) followed by a number (and not Austria, Italy, or Italian within 100 characters);[46] *published by*; *published in*; *publishers*; *, edition*; *edition ,*; *edited by*; *cyclopedia*; *cyclopaedia*; *dictionary*; *his book*; *his article*; *their article*; *his journal*; *quoted from*; *chapter* followed by number; *britannica*; *technical journal*; *handbook*; *chemical society*; *institute of*; *society of*; *proceedings of the*; *bulletin of the*; *textbook*; *scientific american*; *prior patent*; *earlier patent*

Technology classes

We use technological classes to study the impact of libraries on patents in different scientific applications. The USPTO regularly updates its classifications for both new patents and retroactively for older patents. It then publishes these classifications on their website, where we obtain the data. The CUSP uses classifications published in June 2016. We use the CPC (Cooperative Patent Classification) standard, which has eight main subgroups: “Human Necessities”, “Performing Operations; Transporting”, “Chemistry; Metallurgy”, “Textiles; Paper”, “Fixed Constructions”, “Mechanical Engineering; Lighting; Heating; Weapons; Blasting”, “Physics”, and “Electricity.”[47]

[46]We exclude the matches that have Austria, Italy, and Italian within 100 words because they usually refer to citations to patents from those countries. We believe that foreign patent citations are unlikely to be related to materials contained in libraries.

[47]The full taxonomy of classifications can be found at the USPTO website, https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html.