Knowledge Access: The Effects of Carnegie Libraries on Innovation*

Enrico Berkes† Peter Nencka‡

December 2023

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. We study the effect of these libraries on innovation. Patenting in recipient places increased on average by 10–12 percent in the 20 years following library construction relative to a novel control group of cities that applied for but did not build libraries. We show that access to scientific knowledge and increased collaboration opportunities are possible mechanisms.

*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, Bruce Weinberg and three referees 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.

†University of Maryland, Baltimore County, eberkes@umbc.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

There is a widespread consensus that differences in the rate of innovation and technological progress play an essential role in explaining gaps in productivity, economic growth, and intergenerational mobility both across and within countries (e.g., Romer, 1990; Aghion and Howitt, 1992). Studying the inputs of the innovation process and their importance is therefore crucial to understanding growth dynamics. The literature has emphasized the relevance of access to past ideas in fostering 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 innovation by broadening access to knowledge (e.g., United Nations, 2018).

However, estimating the relationship between knowledge-broadening initiatives and subsequent innovation can be challenging. Many institutions that disseminate knowledge do so in narrow ways or target specific groups, such as scientists (e.g., Bryan and Ozcan, 2021). Moreover, institutions that spread knowledge (e.g., colleges) can have other simultaneous effects, like attracting high-skilled immigrants, which makes it difficult to isolate the impact of access to knowledge from other factors that might also affect innovation (e.g., Andrews, 2023).

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 over 1,500 new, high-quality U.S. public libraries financed by steel titan and philanthropist Andrew Carnegie between 1883 and 1919. For residents of recipient cities, 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 access 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 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, economic, and political characteristics. Notably, the two groups of cities also follow parallel patenting trends before library entry.

We find that patenting increases in places that built Carnegie libraries relative to control cities. Patenting starts diverging shortly after receiving a library grant, a pattern 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 10–12 percent in the 20 years after library entry. We show that our findings are not driven by an increase in low-quality patenting or city population. Both women and foreign-born people patented more after libraries opened, suggesting that public libraries helped expand access to knowledge for underrepresented groups, although their relative contribution to patenting remained largely unchanged. We demonstrate that our results are robust to various sample, measurement, and estimation choices. To the best of our knowledge, we are the first to estimate public libraries' effects on innovation.

We investigate two mechanisms that could explain the link between libraries and patenting. 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. Like today, libraries were a central gathering point for meetings and events in the early 20th century. We provide evidence for this social channel by estimating the effect of Carnegie libraries on patents authored by multiple inventors. We find that multi-inventor patenting increased after libraries opened.

Related Literature. This paper relates to three strands of the literature. First, we contribute to the literature on the importance of prior ideas as an input in the production of knowledge. The cumulative and recombinant nature of the knowledge-production process is a cornerstone of recent theories of economic growth (e.g., Weitzman, 1998; Jones, 2021). However, there is much to learn about the empirical relationship between access to knowledge and innovation. Prior work has largely focused on the effects of expanding information access through the patent system or scientific literature, likely impacting people that were already innovators or scientists. For example, a set of papers finds that patent disclosures facilitate future innovation (e.g., Graham and Hegde, 2015; Hegde and Luo, 2018; Gross, 2023) and Furman et al. (2021) finds that patent deposit libraries increased local patenting. Uzzi et al. (2013) find that impactful scientific work combines new ideas with established knowledge. 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 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 that studies 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, 2022; Kantor and Whalley, 2019; Andrews, 2023). These papers consistently find that after the establishment of a college, innovative activity increases, although they tend to 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 and information on the patenting process itself.

Finally, 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 the effects on innovation outcomes and introducing a novel 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. (2023) 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.

2 Historical background

In this section, we describe Carnegie's library program. Appendix C provides additional historical details on the program, libraries before Carnegie's program, and the nature of innovation at the time of Carnegie's grants.

2.1 The Carnegie library program

Andrew Carnegie's library program is one of the most wide-reaching acts of philanthropy in U.S. history. From his first grant in 1883 (to Allegheny, Pa.) 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 believed that public libraries were a way all citizens—including marginalized ones—could improve themselves if they had sufficient drive.

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, expenses, and circulation.

Cities were required to satisfy several requirements to receive construction funds. These requirements, together with other grant features, were made explicit in a letter that Bertram sent to each accepted city, such as 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 construct the library. It also outlines four main requirements of Carnegie's program:[1]

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

[1] 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 restrictions on the specific construction techniques and floor plans that libraries could use, and cities were required to submit blueprints for approval.

charge admission fees.

3. **The construction site needed to be provided by the city.** Carnegie required that the city either purchased a site or re-purposed 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 controversy about the site location. Because libraries often became city 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 ensure that cities could fill the libraries with books, pay the staff, and maintain the building. As illustrated in the above letter, his solution was to require cities to 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 (Bobinski, 1969).

   Despite the written pledge, after granting a library, 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 pledge. In 1917, the Carnegie Corporation—which was founded to manage the library program and related philanthropy—sent a survey to investigate reports that the pledge was not met. The results were stark. In Ohio, 23 out of 77 cities did not meet the pledge (Bobinski, 1969). After discovering this non-compliance, library grant-giving to Ohio was briefly suspended, but there was no direct action against the offending libraries.

2.2 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 buildings. But for some residents, Carnegie grants were controversial. Numerous 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 critical event that generated Carnegie's long-term negative reputation and many eventual rejections was the steelworker 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 could not agree 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 city. The resulting battle led to the deaths of nine strikers, ten militia members, and scores of wounded. The battle made international news. Carnegie's actions were always remembered 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, 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 was a driver of library rejections, it was not always dispositive. For example, Homestead, Pa. itself built a Carnegie library, and the American Federation of Labor president, 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." (Krass, 2011)

Ideological opposition 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. In addition, some cities could not secure or decide on a library site, eventually forfeiting their application.

2.3 Documented relationships between public libraries and innovation

The biographies of inventors provide direct evidence that libraries affected creative and innovative output during this period. These anecdotes do not establish a causal effect of libraries but 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 [sic] 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 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 current and past technical materials, they provided a unique opportunity to discover and recombine existing ideas. Even Thomas Edison regularly used public libraries to access knowledge on electricity and telegraphs that he subsequently used in his inventions (Baldwin, 2001).

Carnegie himself highlighted libraries as a fundamental contribution to his upbringing and education. In his autobiography, he directly tied his interest in library grant-giving to what he learned as a child from the private library of Colonel James Anderson:

> It was from my own early experience that I decided there was no use to which money could be applied so productive of good to boys and girls who have good within them and ability and ambition to develop it, as the founding of a public library in a community which is willing to support it as a municipal institution (Carnegie, 1920).

3 Data

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. We assign each library to the city in which it was built and use grant years as "treatment" years since they are available both for cities that accept and reject grants. In some larger cities, Carnegie funded multiple libraries. These awards typically paid for the construction of the main library building and branch libraries. Multi-library grants occurred in approximately 5 percent of recipient places, accounting 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.[2] As described in Appendix C, we also collect information on when each library opened.

To identify cities that rejected Carnegie library grants once they were approved, we rely on Bobinski (1969). Bobinski (1969) identified 209 Carnegie libraries "that never materialized." His

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

primary source is the original Carnegie library correspondence between Bertram and rejecting cities.[3] In addition to the locations of the rejecting cities—and, when possible, the reason for rejection—Bobinski (1969) identifies the grant amount and the date of the offer.

From the universe of library grants, we construct a consistent sample of cities that we use throughout the empirical analysis. First, we exclude eight grants that Carnegie donated before 1899 during the so-called "retail period." Carnegie hand-selected the location of the grant recipients before opening his program to national applications in 1899, raising selection concerns. Second, we exclude Carnegie grants in more populous cities and counties. Carnegie grants in larger cities were distinct and came with additional requirements and benefits relative to the other grants. Moreover, many of these cities and counties (e.g., Cook County, Illinois) already had well-developed library systems that Carnegie improved. Our baseline sample is constructed by excluding all cities with more than 25,000 people and all counties with more than 750,000 people in 1900.[4] Third, we exclude control cities that rejected Carnegie because they instead built a library with funds from a local philanthropist. Finally, we exclude 231 cities that cannot be uniquely matched to the 1900 census, our source of pre-program city characteristics. These cities were often unincorporated in 1900, and census enumerators did not record their names; this was particularly common in the Western United States. We show in Section 6 that these restrictions do not affect our main conclusions.[5]

[3] Bobinski (1969) used newspaper articles and surveys sent out to cities that applied to the Carnegie program as additional sources.

[4] We show that our results are robust to stricter alternative cutoffs and including additional cities in Section 6.

[5] A complete description of how each sample restriction affects our sample size and estimated patenting impacts are shown in Table A9.

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 crucial 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). We describe this dataset in more detail in Appendix C.

3.3 City and county covariates

For our empirical analysis, we collect city's time-varying population from Steiner (2018).[6] For other 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 complement our data with information on union activity (Bittarello, 2019), presidential elections (Gentzkow et al., 2011), local newspapers (Gentzkow et al., 2011), and college locations drawn from 1900-1901 Bureau of Education Directory. We describe these data and their sources in more detail in Appendix C.

3.4 Summary statistics

Figure A1 shows the cumulative distribution of grant dates for cities that built (panel A) and rejected (panel B) Carnegie libraries. Both panels display similar distributions. Very few libraries

[6] Several 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 population estimates over time. We interpolate population between decennial censuses.

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 A2 shows the geographical distribution of places in our sample. Blue dots represent places that built a library, while red dots indicate places in the control group.[7] As this map illustrates, the reach of Carnegie's program was national. Each state, except Alaska, Delaware, and Rhode Island, received at least one Carnegie library. However, some geographical patterns are apparent. Granted libraries were popular in the Midwest and Northern states.[8] Indiana received the most Carnegie libraries, with 164 built in 157 cities. As with cities that received a library, rejecting cities are located across the United States, though some interesting geographic patterns across regions emerge.

For example, cities in Southern states were more likely to reject libraries. Most libraries in the South were segregated during this period. Carnegie did not require libraries in the South to integrate in order to receive funds, nor did he require that cities building a "whites-only" library also build a branch for Blacks. But there was some local and legal pressure for "separate-but-equal" library access, which might have contributed to the South's propensity to reject libraries. For these reasons, in the rest of the paper, we confine most of our discussion to within-state comparisons of accepting and rejecting cities. We also show that our results are robust to excluding the South from our analysis.

Table A1 summarizes the variables that we use in our analyses. Panel A of Figure 1 compares 1900 covariates across cities that received and those that rejected Carnegie libraries. Although our empirical strategy does not require it, a balanced sample alleviates concerns regarding systematic

[7]Figure A3 provides a visual representation of the expansion of the Carnegie program over time.

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

differences across city characteristics that might correlate with library decisions and patenting outcomes. We plot the coefficient from an indicator variable for building (rather than rejecting) a Carnegie library from separate regressions on standardized covariates conditional on state and grant-year fixed effects. The covariates include basic demographics, population, the share of children in school, the share of adults in various occupation and industry groups, and the number of local Knights of Labor union branches and strikers.[9] These results suggest that places that accepted and rejected Carnegie libraries were broadly similar on many observable characteristics before Carnegie's program. Panel B of Figure 1 shows a starkly different result when we compare cities that built Carnegie libraries and those that did not apply for a grant. Particularly, non-applicants have fewer workers in plausibly high-innovation occupations—such as craftsmen—and have lower imputed earnings, among other differences.

4 Empirical analysis

4.1 Patenting trends

In Figure 2, we plot city-level aggregate patenting for our treatment and control samples twenty years before and thirty years after library grants. In panel A, we show the inverse hyperbolic sine of patent counts over time conditional on grant-year fixed effects; we show the same figure for patent counts in panel B. Cities that built and rejected Carnegie libraries follow parallel patenting trends before libraries were granted. The trends start diverging shortly after grant years, consistent with construction times. The dashed red line indicates the average number of years from when the grant was received to the library's opening. Patenting differences between

[9] 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.

cities that accepted and rejected libraries peak between 5 and 20 years after library receipt.[10] This difference declines over time, and 30 years after library acceptance, cities that accepted and rejected libraries revert to similar levels of patenting.

4.2 Difference-in-differences estimates

To formalize these patterns, test how our results change when we control for covariates, and explore statistical uncertainty, we estimate difference-in-differences regression models of the form:

$$PatentMeasure_{i,s,t} = \beta_1 Post_{i,t} + \beta_2 Library_i \times Post_{i,t} + \delta_{s,t} + \lambda_{g,t} + \gamma_i + \epsilon_{i,s,t} \tag{1}$$

where $PatentMeasure_{i,s,t}$ is a measure of patenting activity 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 library grant approval; it is a function of both time (t) and the grant year of city i. Note that $Post_{i,t}$ is well-defined for all units in treatment and control groups since all cities were offered library grants. The baseline empirical model includes state-by-year ($\delta_{s,t}$), grant year-by-year ($\lambda_{g,t}$), and city (γ_i) fixed effects. The coefficient of interest, β_2, identifies the average patent increase after receiving a grant in cities that built a library relative to cities that were offered a grant but did not build a library. We estimate our model using observations 20 years before and after library grants for each city and cluster standard errors by cities.

Here, it is useful to relate Equation 1 to recent concerns in the applied econometrics literature about interpreting staggered difference-in-differences estimators. For example, Goodman-Bacon

[10] Figure A4 displays the distribution of time required to build a library after grant receipt. The pattern of treatment effect growth aligns with the timing of results that Furman et al. (2021) find for the effects of patent deposit libraries.

(2021) and De Chaisemartin and d'Haultfoeuille (2020) show that the two-way fixed effect (TWFE) estimator may be biased. In particular, the TWFE approach in part uses units treated in earlier periods as controls for later-treated units. If treatment effects are dynamic (e.g., grow or shrink as a function of time), these comparisons will likely produce biased estimates of the average treatment effect (Goodman-Bacon, 2021). In contrast to the standard TWFE case when control group treatment timing is not observed, in our empirical framework, we observe grant years for both types of units. This allows us to directly compare pre-post changes across treatment and control groups and avoid using the post-period of some of our treatment units as a "control" for other treated units. In addition, by including grant-by-calendar year fixed effects in our baseline specification, we exploit variation in treatment status among groups of cities that received library grants in the same year. This approach is similar to "stacked" regression estimators introduced recently. We show additional robustness to new difference-in-differences methodologies in Section 6.

Our main results are reported in Table 1. We show eight specifications, each corresponding to combinations of different fixed effects and covariates. Estimates for the key coefficient (β_2) suggest that patents counts increased by approximately 10–12 percent in the years after the receipt of library grants in cities that built libraries relative to cities that did not build libraries.[11] 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 when we exclude the Great Depression, which led to a sharp decrease in

[11]Under-powered analyses can be prone to magnitude and even sign errors. In Appendix D, we show that our analysis is well-powered to detect effects of the magnitude that we estimate using a simulated power exercise as proposed by Black et al. (2022).

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

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

where the terms in Equation 2 are defined similarly as in Equation 1. Instead of a single post-period indicator, we interact treatment status with dummy variables for years relative to Carnegie grants ($RelYear_{i,t,r}$). The vector of coefficients $\beta_{2,r}$ traces 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 before and after the treatment date, this specification can also be used to assess the plausibility of the parallel trends assumption. 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 3 plots the marginal effects of receiving rather than rejecting a library over time, binning relative years in 5-year increments. We cluster standard errors at the city level and plot 90 percent confidence intervals. The results in Figure 3 are consistent with the dynamics outlined by the raw data plotted in Figure 2. Flat pre-trends indicate that cities that received and rejected libraries follow similar patenting patterns before library grants. Patenting behavior diverges quickly after receiving a library grant, consistent with data on construction time. Figure 3 also shows that these patterns are consistent across different outcome variable transformations. Whether we use inverse hyperbolic sine patenting, log patents plus 1, patent counts, a Poisson estimation of patent counts, or condition on observing at least one patent in each city-year, we observe similar pre- and post-period patterns.

Figure 4 shows the same results without year binning. This figure shows similar, albeit less precisely estimated, patterns.[12] Taken together, these results suggest that Carnegie libraries increased innovation, measured through patenting activity, but that these increases were not permanent and started disappearing after about 20 years.

4.3 Identification

To interpret our findings as causal, the standard difference-in-differences assumption must hold: conditional on included covariates, patenting in cities that did not build their granted library would have followed the same path as in cities that did build their libraries.

Given the absence of pre-trends, it is unlikely that we are picking up trend differences that increase the likelihood of building a library and patenting simultaneously. If that were the case, we would expect differences in patenting trends even before libraries opened. Nevertheless, contemporaneous changes in either our treatment or control cities could correlate with our treatment and have an independent effect on patenting. Proving the lack of these shocks is ultimately untestable, but it is possible to provide suggestive evidence that other things are not changing simultaneously with library entry.

First, Appendix Table A2 shows no effect of library expansion on the number, existence, or partisan orientation of local newspapers. Table A2 also shows no time-varying correlation between library entry and county-level presidential vote shares. These results suggest that the library treatment is not simply picking up local changes in news organizations or political pref-

[12] Figures A9 and A10 show that we obtain similar patterns of results when we condition on only state-year and city fixed effects across transformations.

erences.

Second, our results are robust to a rich set of 1900 controls interacted with grant-year-by-year fixed effects, including local occupation and industry distributions, the share of children in school, the presence of a college, basic demographics (e.g., share of Black people, average age), and union activity. As shown in Table A3, in all these specifications, we find estimates between 10 and 13 percent, close to our baseline range. In addition, we obtain similar estimates when we include time-varying controls for population or consider per-capita patenting as our outcome variable. This result suggests that population growth, which can proxy for many time-varying characteristics that make a city attractive, cannot alone explain our results.

4.4 Additional Results

4.4.1 Patent Quality

Not all patents have the same innovative content; even among the most innovative patents, real-life impacts and value can differ significantly. If libraries only increased lower-quality patenting, the impact discussed in the previous sections would be overstated. To explore this possibility, we consider three different but related measures of quality: (1) the probability that a city-year observation had a patent cited in the future by other patents, (2) the count of future patent citations, (3) the probability that a city-year produced a patent in the top ten percentile of the text-based quality measure proposed by Kelly et al. (2021). Citations measure whether a patent has 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 A4 shows evidence for a slight increase in patent quality when we consider the two measures based on patent citations and a null effect according to the text-based measure.[13] These findings suggest that changes in patent quality do not drive the results described in Section 4.2. Libraries appear to have increased the patenting rate without a decrease in quality. If anything, they slightly increased it—as measured by citations.

4.4.2 Spillovers to nearby cities

Our main analysis compares 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 places.[14] 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 directly impacted innovation in nearby areas, and our estimates would understate the actual impact of libraries. On the contrary, our estimates would overstate the true effect if innovators moved to or filed patents in cities with libraries instead of filing in their hometowns.

To test for spillover effects, we construct a set of 'doughnut' treatment and control groups that incorporate 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 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

[13] The opening of a library increases the likelihood of observing at least one patent cited in a city-year observation by 5.1 percentage points (mean 29.8%) and the total number of citations by 1.3 citations (mean 2.7).

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

grant dates based on the closest accepted or rejected library and cluster standard errors at the city-of-grant level.

Figures A5 and A6 report our estimates when we consider nested and non-nested treatment and control groups, respectively. We find evidence of spillover effects within 15 miles, with the patenting activity of places within this radius increasing by about 8% compared to the control group. These effects quickly decrease to zero as distance increases. Thus, our findings do not appear 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.

Importantly, this spillover test does not estimate potential effects due to migration from locations that are not nearby. Migration is an interesting possible mechanism: we can imagine that libraries attracted those most interested in access to knowledge, an interest plausibly correlated with patenting. It is difficult to quantify this mechanism's size since it is challenging to identify migrants in the patent data. That said, as previously discussed in Section 4, our results are robust to controlling for time-varying population and using per-capita patenting as an outcome, implying that our findings cannot be explained entirely by aggregate migration patterns.

4.4.3 Heterogeneity

Since libraries were open to all, they might have benefited those who had less access to other information and education sources at this time in history. These groups include women and immigrants, both 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 essential for new inventors since those with more experience might already have access to alternative knowledge sources—although the qualitative evidence that we reviewed suggests that established inventors also heavily relied on libraries.

We find evidence that libraries boosted innovation produced by immigrant and female inventors (Table A5). In our baseline specification, the opening of a Carnegie library increases the number of inventions patented by women and immigrants by 0.02 and 0.09 (relative to means of 0.09 and 0.21), respectively. However, in both cases, we do not find evidence that the *shares* of inventors in these categories increased, likely because libraries also impacted non-immigrant and non-women inventors. Similarly, we see limited evidence of an effect on the extensive margin of patenting at either the city (Table A6) or individual (Table A7) levels. While first-time inventors increase, their share stays roughly the same after libraries opened.

Beyond inventor characteristics, we also explore the characteristics of *cities* that might have changed the effectiveness of libraries. Overall, we find imprecisely estimated but positive complementarities between libraries, the share of youth in school, the share of craftsmen, and the city's size. While our sample size does not give enough statistical power to detect small heterogeneous effects, our results, reported in Table A8, suggest that the effects of libraries did not dramatically differ across the city characteristics considered here.

5 Mechanisms

In this section, we explore 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 explore the first potential mechanism by observing that if an information channel were indeed at play, not all patent classes would be affected similarly. In fact, we would expect that technology classes closer to the type of knowledge contained in libraries would be affected disproportionately more. To explore this information link between library holdings and the increase in patenting activity that we observe, we collect data on the types of books that libraries carried during this period.

In 1904 the ALA published a list of about eight thousand suggested books for new libraries. To classify them, we rely on the Dewey decimal numbers associated with each volume to identify technical books and the topics they cover. More precisely, we focus on books with numbers in the 500 (Natural Science) and 600 (Applied Science) categories and divide them further based on the second digit of their classification. Figure A7 shows the distribution of these books across the 20 resulting categories. The ALA's list included a subset of about 900 books related to science and technology. Many of them were practical "how-to" books related to, for example, Engineering and Zoological Sciences. Intuitively, it seems plausible that public libraries would have a minor 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 only available in research libraries and universities. Moreover, the most technical fields likely required access to expensive machinery outside the reach of many citizens.

Although the ALA's list provides valuable insight into the type of books that prominent librarians considered to be most valuable at the beginning of the 20th century, it is unclear how influential this list was and which books each library—especially smaller ones—decided to include in their collection. In order to shed light on this, we use the pre-1950 book catalog of five libraries made available by the Main Street Public Library Database (Wiegand, 2011).[15]

[15]These are the public libraries of Lexington, Mich.; Morris, Ill.; Osage, Iowa; Rhinelander, Wis.; and Sauk Center, Minn.. While only one of these libraries was a Carnegie library (Rhinelander, Wis.), we believe they provide a

Figure A8 shows the distribution of books across the 20 categories based on the Dewey decimal numbers for the five libraries separately and combined.[16] The bar charts show that libraries tended to carry more books on specific topics, including Science, Zoological Sciences, Medical Sciences, Engineering, Agriculture, and Domestic Sciences. Science is a general category that includes books such as encyclopedias and collections of magazines such as *Scientific American*.[17] By contrast, the libraries had fewer books in less mechanical science areas, such as Physics and Chemistry.

This distribution of topics closely mirrors the technology classes that are most affected by the arrival of a library. To show this, 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 2 suggest that libraries had the largest effects on patenting in classes corresponding to the practical trades, such as Construction, Transport, and Mechanical Engineering. On the other hand, we find no effects in Chemistry and Physics: the corresponding coefficients are small and statistically indistinguishable from zero.

To further explore this mechanism, we test whether we observe more patents that cite prior materials in cities that received a Carnegie library compared to control cities. If patrons used materials in libraries for their inventions, we might expect the number of these citations to increase. We identify patents that cite prior materials by identifying a set of keywords associated with a

representative sample of the books found in small local public libraries during our study period.

[16]Note that Dewey decimal numbers are not available for all the volumes in the database.

[17]During this period, *Scientific American* reported a list of the patents granted in the previous month together with a short description of the inventions deemed most relevant.

citation in a training dataset and searching the corpus of remaining patents for these keywords.[18]

For this analysis, we estimate regressions analogous to Equation 1 with an outcome variable that indicates whether a patent in a particular city-year cited a book, previous patent, or magazine. In Table 3, we show that patents that cite past materials increased more in cities that built rather than did not build Carnegie libraries. In particular, building a library is associated with approximately a 0.3 to 0.8 percentage point increase in the probability of observing a patent that cites prior materials, though the effects are imprecisely estimated in some specifications. The largest and most precisely estimated effects arise when we focus on the pre-1925 sample.[19] This effect is large relative to the baseline probability of observing prior-work-citing patents (1.1 percent). Table 3 also shows that the count of patents that cite prior work increases after libraries are built, though as with the probability results, the precision increases when we focus on earlier sample periods.

Knowledge is available in more than just books and other physical media. It is also accessed through informal interactions with other people. Such exchanges may increase creative output—for example, Andrews (2020) finds that reduced collaboration after prohibition reduced aggregate patenting. For this reason, we examine whether libraries also affect collaboration. Community-centric programming was standard at libraries in the early 20th century. Carnegie's suggested library blueprints included meeting rooms for community activities, and historical accounts of the Carnegie program highlight the variety and number of groups and clubs that used libraries

[18]More details on this procedure are given in Appendix C.

[19]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.

for their meetings and activities (Learned, 1924). This programming suggests that libraries might have been places fostering networks of innovators.

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 might 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 analogs of our baseline difference-in-differences regression with an indicator for observing a multi-authored patent as the outcome variable. Table 3 shows the results from this analysis. Building a library is associated with approximately a 2 percentage point increase in the probability of observing a patent that cites prior materials, though the effects are imprecisely estimated in some specifications. Table 3 also shows that the number of multi-author patents increases after libraries are built.

6 Robustness of patenting results

Table A9 shows the results of sample-based robustness checks. The first panel reports the estimates from Equation 1 for subsets of the baseline sample. First, since the South was generally slower to build public infrastructure during this 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 several control cities whose primary reason for rejecting the grant was related to financial considerations according to Bobinski (1969). In particular, these cities may have been worried about Carnegie's requirement that cities pledge 10 percent of the cost of the building on an annual basis for maintenance. Finally, we 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 smaller places and rural areas. Estimated coefficients are modestly smaller when we focus on cities with less than 5,000 people, consistent with the heterogeneity results described in Section 4.4.3.

The second panel of Table A9 shows a similar set of exercises where we 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, cities missing 1900 covariates, and cities that we cannot uniquely match to the census. 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 A10, we show a series of robustness checks using alternative patent measures, including $ln + 1$ transformations and patent counts. Table A10 shows that when we use these alternative measures, the results are positive and, in percentage terms, consistent with or larger than our baseline estimates.[20] We also show estimates obtained from the Poisson count model. In each case, results are comparable or larger in percentage terms relative to our baseline estimates. Figure A11 shows a version of the yearly results where we use library opening dates (instead of grant dates) as our defined treatment start period—the results are again similar.

Next, we estimate aggregate versions of our baseline regression models. In particular, we sum patent counts for each city before and after library grants are made. We report models with city fixed effects, city and state-post period fixed effects, and city and grant-year-post fixed effects for different transformations of patent counts and different estimators. Results are shown in Table

[20]Tables A11 and A12 report the full set of specifications for the $ln(patents + 1)$ transformation and patent counts, respectively. These tables show a similar pattern of results as our main specification.

A13 and are similar to or larger than our baseline estimates.

Lastly, we include additional robustness checks based on our model specification. Because we have treatment timing for both our treatment and control units, our setting is somewhat unique. An alternative approach involves not using this timing for the control group. In this case, our empirical framework falls into the category of a staggered difference-in-differences approach. We have implemented two of the newly proposed estimators for this setting. First, we use the Gardner (2021) imputation estimator.[21] We show the results of this estimation in Figure A12 for our state-year model. We see a similar pattern of results to our baseline estimator.

Second, we estimate a "traditional" stacked estimator. In particular, we stack each treatment cohort along with all rejecting cities and estimate our baseline model. Figure A13 shows the results of this exercise. We interact all fixed effects with "cohort" fixed effects, defined as the relevant treatment year g. We show these results using an inverse hyperbolic sine transformation, conditioning on non-zero patents, using untransformed patent counts, and estimating a Poisson model of patent counts. In all cases, we observe results similar to our baseline estimates. Similarly, Figure A14 shows that we get similar results when we limit the stacked control cohorts to +1/-1 year from the treatment grant year.

7 Conclusion

This paper provides new empirical evidence for the importance of existing ideas as an input in the knowledge production process. Leveraging the rollout of public libraries promoted by Andrew Carnegie's grants in the late 19th and early 20th centuries, we test whether cities that

[21]We calculate standard errors using a Bayesian clustered bootstrap, drawing independent exponential weights for each cluster (city) in our panel over 1,000 iterations.

built libraries experienced an increase in innovative activities. We find that patenting increased by 10–12 percent in cities that built libraries relative to a 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 innovation becomes more complex and team-based. Our results are robust to various sample, measurement, and estimation choices.

Our setting provides a unique opportunity to test the hypothesis that information-providing institutions impact innovation. Unlike many 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 a meaningful impact on innovation. 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 similar 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 libraries' historical and contemporary effects. Our treatment and control groups are drawn from a set of similar cities that all applied for libraries. It is therefore not possible to extrapolate our results to the entire United States if there are important interactions between city characteristics and the importance of libraries. Exploring these interactions is a crucial avenue for future work. While the emergence of the internet has diminished the unique informational role of public libraries, 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 we documented in the early 20th century.

References

Aghion, P., L. Boustan, C. Hoxby, and J. Vandenbussche (2009). The causal impact of education on economic growth: Evidence from US. *Brookings Papers on Economic Activity 1*(1), 1–73.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

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. J. (2020). Bar talk: Informal social interactions, alcohol prohibition, and invention.

Andrews, M. J. (2021). Historical patent data: A practitioner's guide. *Journal of Economics & Management Strategy 30*(2), 368–397.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? Evidence from the establishment of US colleges. *American Economic Journal: Economic Policy*. Forthcoming.

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). 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 13*(4), 218–260.

Bittarello, L. (2019). Organizing collective action: Labor strife in the US in the 1880s.

Black, B., A. Hollingsworth, L. Nunes, and K. Simon (2022). Simulated power analyses for observational studies: An application to the Affordable Care Act Medicaid expansion. *Journal of Public Economics 213*, 104713.

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 spillovers.

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.

Carnegie, A. (1920). *The Autobiography of Andrew Carnegie*. Public Affairs.

Currie, J. and J. Ferrie (2000). The law and labor strife in the United States, 1881–1894. *The Journal of Economic History 60*(1), 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.

Gardner, J. (2021). Two-stage differences in differences.

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

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review 101*(7), 2980–3018.

Gilpin, G., E. Karger, and P. Nencka (2023). The returns to public library investment.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277.

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

Gross, D. P. (2023). The hidden costs of securing innovation: The manifold impacts of compulsory invention secrecy. *Management Science 69*(4), 2318–2338.

Hausman, N. (2022). University innovation and local economic growth. *The Review of Economics and Statistics 104*(4), 718–735.

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, C. I. (2021). Recipes and economic growth: A combinatorial march down an exponential tail. Working Paper 28340, National Bureau of Economic Research.

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 3*(3), 303–320.

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). Localized knowledge spillovers and patent citations: A distance-based approach. *The Review of Economics and Statistics 96*(5), 967–985.

Researching NYC (2015). Chester F. Carlson. https://researchnychistory.wordpress.com/2015/06/06/1361/.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy 98*(5), 71–102.

Rosenbloom, J. L. (1998). Strikebreaking and the labor market in the United States, 1881–1894. *The Journal of Economic History 58*(1), 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.

Steiner, E. (2018). Spatial History Project, Center for Spatial and Textual Analysis, Stanford University. https://github.com/cestastanford/historical-us-city-populations.

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. https://www.un.org/sustainabledevelopment/blog/2018/01/worlds-vulnerable-countries-track-achieve-universal-internet-access-2020-un-report/.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013). Atypical combinations and scientific impact. *Science 342*(6157), 468–472.

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

Wenyon, M. (2009). NYPL, Mother of Invention. https://www.nypl.org/blog/2009/01/30/nypl-mother-invention.

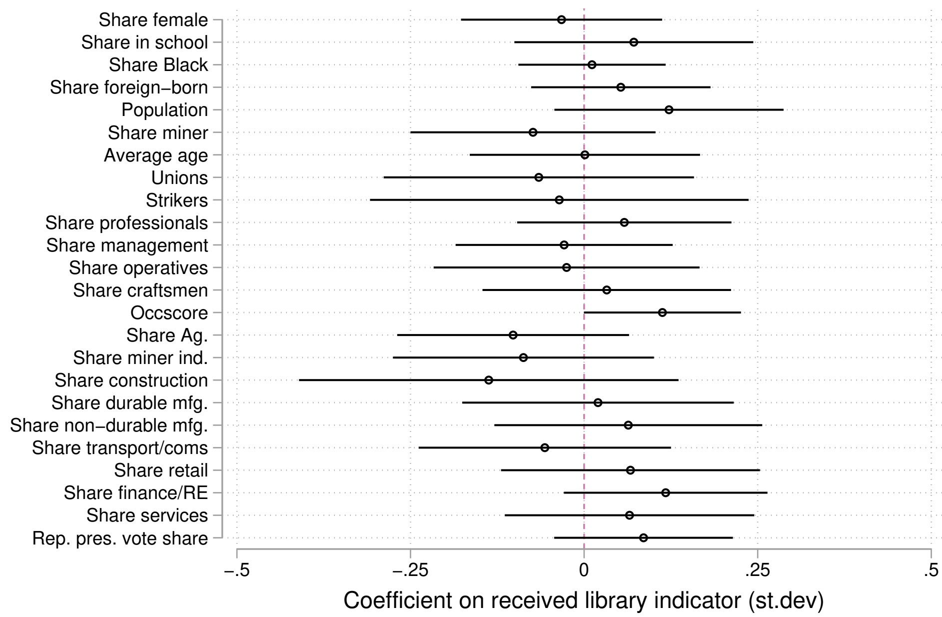
Wiegand, W. (2011). Main Street Public Library Database. https://cardinalscholar.bsu.edu/handle/123456789/194598.

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

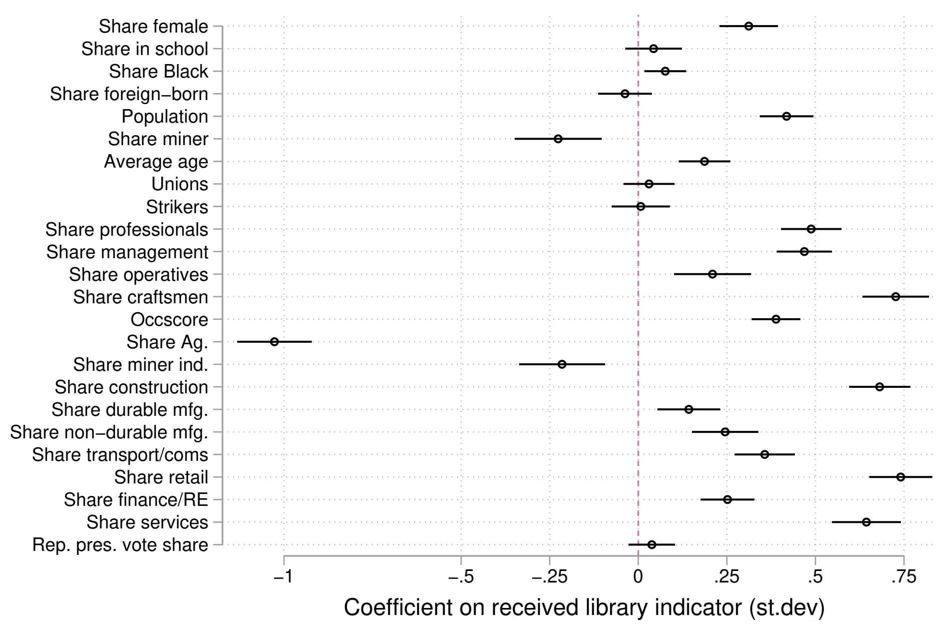
Xu, X., A. Watts, and M. Reed (2019). Does access to internet promote innovation? A look at the US broadband industry. *Growth and Change 50*(4), 1423–1440.

Figure 1: **Comparison of city characteristics between Carnegie and rejecting cities and Carnegie and non-applicant cities**

(a) Carnegie vs. rejecting cities



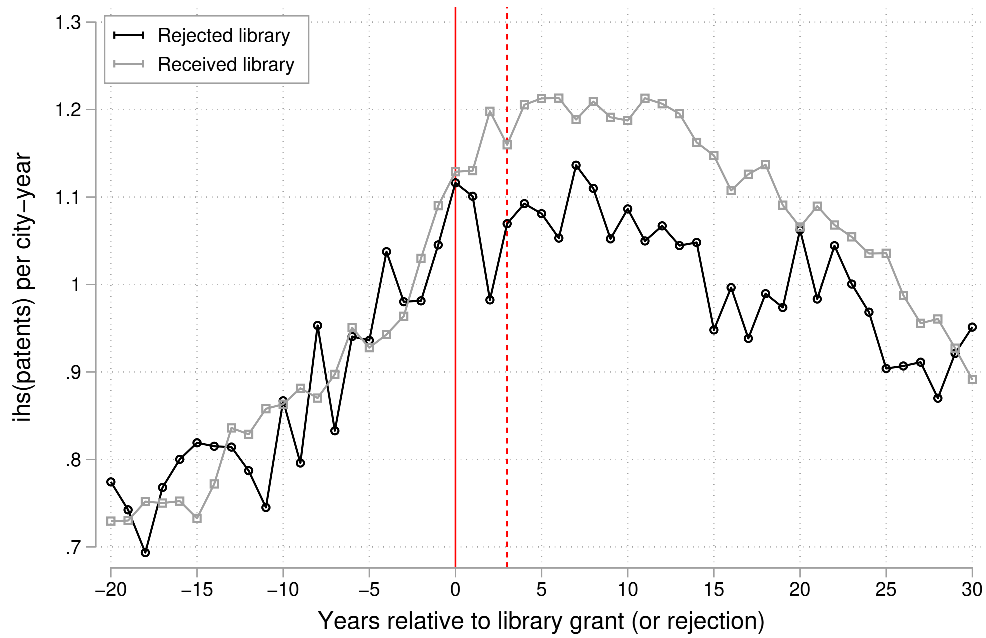
(b) Carnegie vs. non-applicant cities



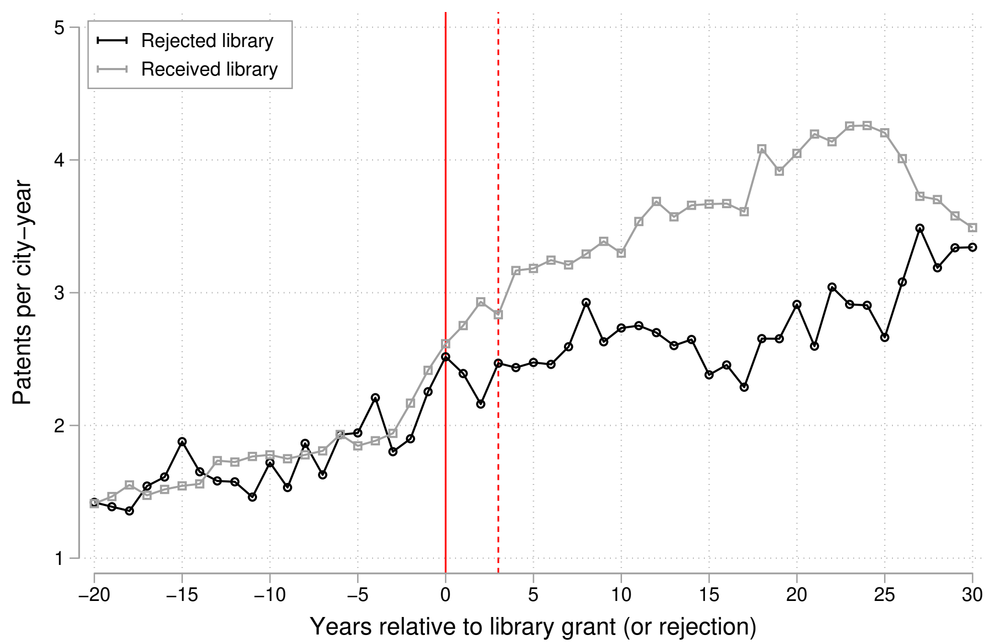
Notes: These figures compare city characteristics between cities that received Carnegie libraries and those that rejected them (Panel A), as well as Carnegie libraries and 4,090 cities with over 1,000 people that did not apply to the Carnegie program (Panel B). We plot the standardized coefficient from separate regressions of each variable on a dummy equal to 1 for a city that built a library conditional on state and grant year fixed effects. The median grant year of Carnegie cities in their state is used for never-applicants. Covariates are standardized to have a mean of zero and a standard deviation of 1. We report 95% confidence intervals. Covariates, measured in 1900, are all at the city-level except for the Republican Presidential election vote share, which is county-level. In this case, each town in our sample within the same county is assigned the same vote share.

Figure 2: **Average patenting activity in treatment and control groups relative to grant years**

(a) Inverse hyperbolic sine of patent counts



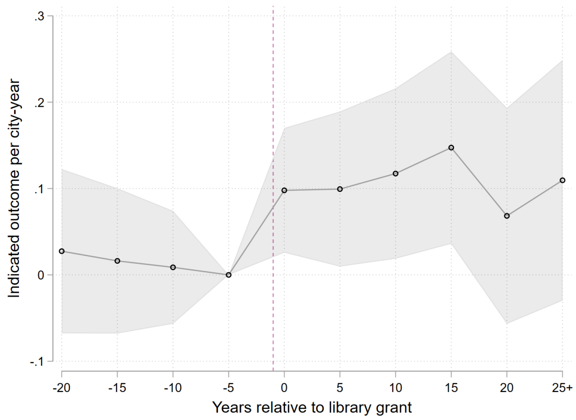
(b) Patent counts



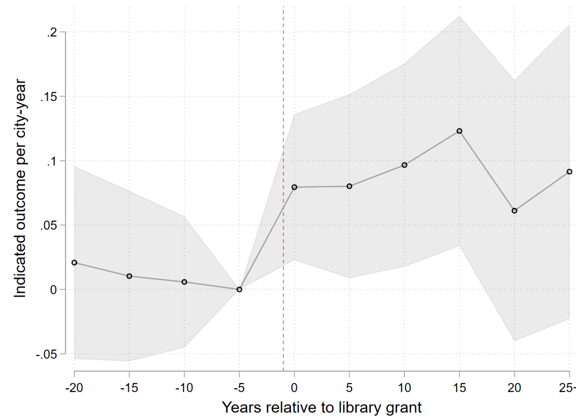
Notes: This figure shows average patenting in cities that received a Carnegie library and cities that were approved for a Carnegie grant but did not build a library after controlling for grant-year fixed effects. Panel A shows $ihs(patents)$ and Panel B shows untransformed patent counts. Averages are relative to grant dates. The solid vertical line indicates library grant dates. The dashed line indicates the mean time from library grants to opening. The average opening time was three years.

Figure 3: **Event study estimates of Carnegie libraries on patenting for alternative patent measures**

(a) Inverse hyperbolic sine of patent counts

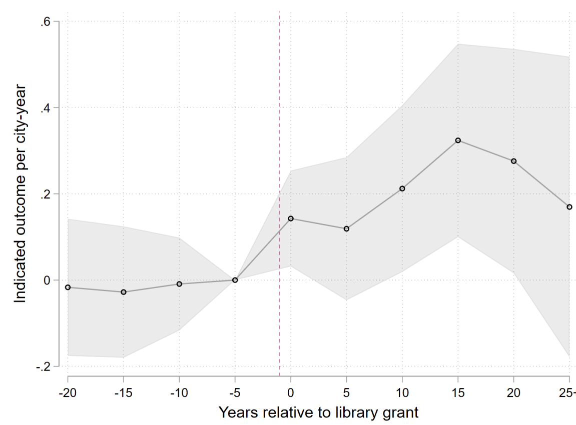


(b) Log patents + 1



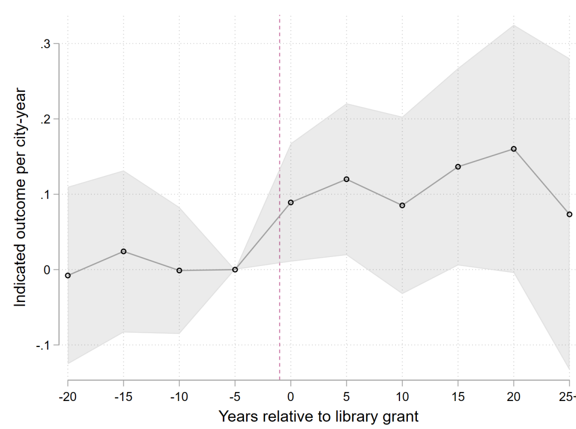
(c) Patent counts

(d) Patent counts (Poisson)



(e) Inverse hyperbolic sine, $pat > 0$

(f) Log patents, $pat > 0$

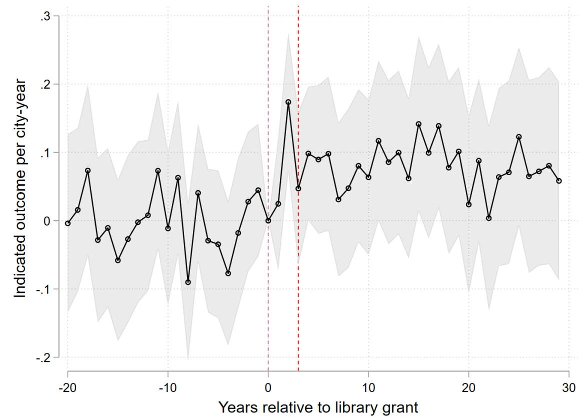


Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. Coefficients are binned in 5-year increments. All models include state-year, city, and grant-year-by-year fixed effects. Standard errors are clustered by city.

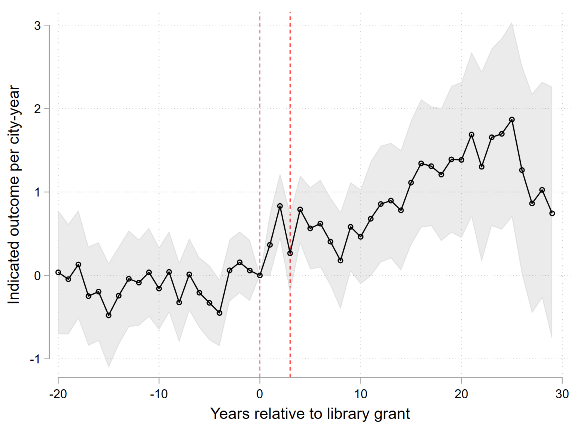
Figure 4: **Event study estimates of Carnegie libraries on patenting for alternative patent measures (yearly)**

(a) Inverse hyperbolic sine patents

(b) Log patents + 1

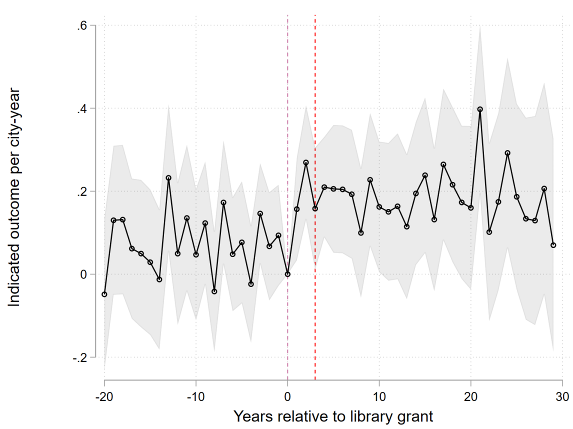


(c) Patent counts

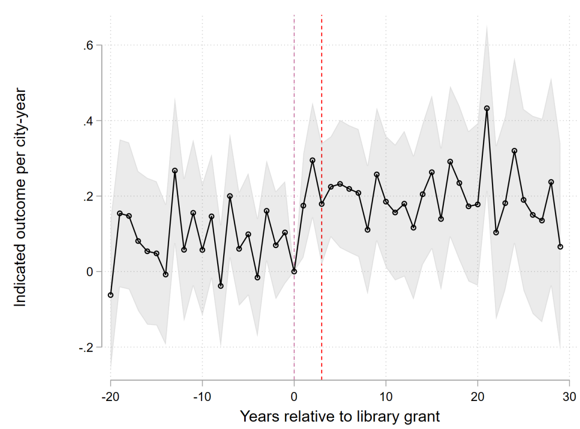


(d) Patent counts (Poisson)

(e) Inverse hyperbolic sine, $pat > 0$



(f) Log patents, $pat > 0$



Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. All models include state-year, city, and grant-year-by-year fixed effects. Standard errors are clustered by city.

Table 1: Effect of Carnegie libraries on patenting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.120	0.118	0.113	0.112	0.112	0.142	0.129	0.105
	(0.049)	(0.049)	(0.047)	(0.047)	(0.047)	(0.051)	(0.050)	(0.050)
Pre-1929 sample								
Built library × post	0.154	0.144	0.137	0.133	0.117	0.171	0.136	0.111
	(0.049)	(0.049)	(0.048)	(0.048)	(0.047)	(0.053)	(0.049)	(0.050)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean ihs(pat)	1.007	1.007	1.007	1.007	1.007	1.007	1.007	1.007
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

Notes: This table shows the estimated impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. We report the estimates obtained 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 is equal to 1 in the years after cities was approved for a 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, we use the earliest grant year. The outcome variable is the inverse hyperbolic sine of patents. Each observation is at the city-year level. Standard errors are shown in parentheses and clustered by city.

Table 2: Difference-in-differences estimates by patent classes

Patent class	S-Y FE		S-Y and GY-Y FE	
	Lib. × post	Std. err.	Lib. × post	Std. err.
Human necessities	0.060	(0.025)	0.043	(0.026)
Performing ops/transport	0.117	(0.031)	0.094	(0.032)
Chemistry	0.022	(0.026)	0.013	(0.025)
Textiles	0.007	(0.015)	0.002	(0.014)
Constructions	0.048	(0.018)	0.030	(0.018)
Mech. engineering	0.094	(0.026)	0.071	(0.026)
Physics	0.022	(0.019)	0.008	(0.019)
Electricity	0.050	(0.018)	0.042	(0.019)

Notes: Each row represents a separate estimate of the model in Equation 1 using as an outcome variable the inverse hyperbolic sine of patents that fall in the indicated technology class according to the Cooperative Patent Classification. Models include city and state-year (first columns), and state-year and grant-year-by-year (second columns) fixed effects. Standard errors are clustered by city.

Table 3: Effect of Carnegie libraries on patents that cite past work and coauthored patents

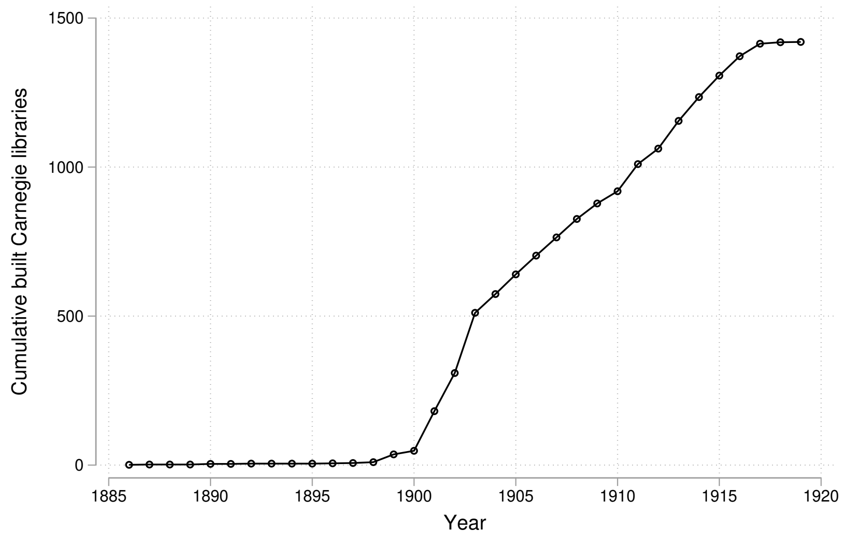
Outcome and sample	S-Y FE		S-Y and GY-Y FE	
	Lib. × post	Std. err.	Lib. × post	Std. err.
Observe a work-citing patent				
Full sample	0.005	(0.004)	0.003	(0.004)
Pre-1929 sample	0.006	(0.004)	0.002	(0.003)
Pre-1925 sample	0.008	(0.004)	0.003	(0.003)
# work-citing patents				
Full sample	0.009	(0.005)	0.007	(0.006)
Pre-1929 sample	0.010	(0.005)	0.008	(0.005)
Pre-1925 sample	0.011	(0.005)	0.008	(0.005)
Observe a multi-author patent				
Full sample	0.022	(0.015)	0.016	(0.015)
Pre-1929 sample	0.024	(0.015)	0.017	(0.015)
Pre-1925 sample	0.027	(0.016)	0.019	(0.016)
# multi-author patents				
Full sample	0.125	(0.043)	0.099	(0.044)
Pre-1929 sample	0.113	(0.037)	0.083	(0.037)
Pre-1925 sample	0.111	(0.037)	0.077	(0.036)

Notes: The upper panel of this table shows the effect of Carnegie libraries on the probability of observing a prior-work citing patent and their count. The bottom panel reports estimates of the effect of libraries on the probability of observing a multi-author patent and the count of such patents. Models include city and state-year (first two columns) and state-year and grant-year-by-year (second two columns) fixed effects. Standard errors are clustered by city.

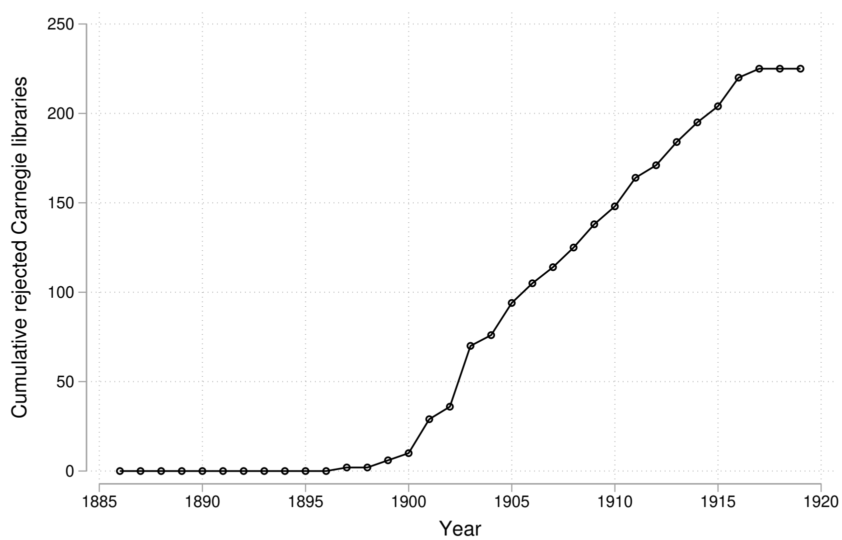
A Figures and tables for online publication

Figure A1: **Cumulative distribution for cities that built a Carnegie library and cities that rejected a Carnegie grant by grant-year**

(a) Cities that built Carnegie libraries

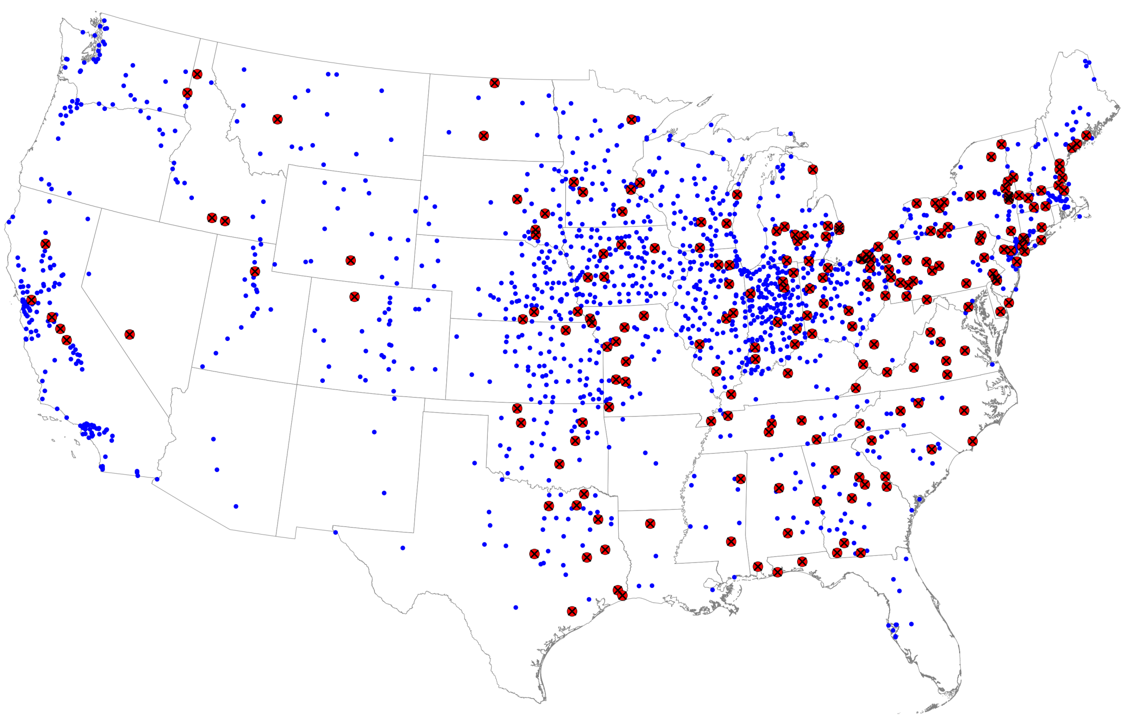


(b) Cities that rejected Carnegie libraries



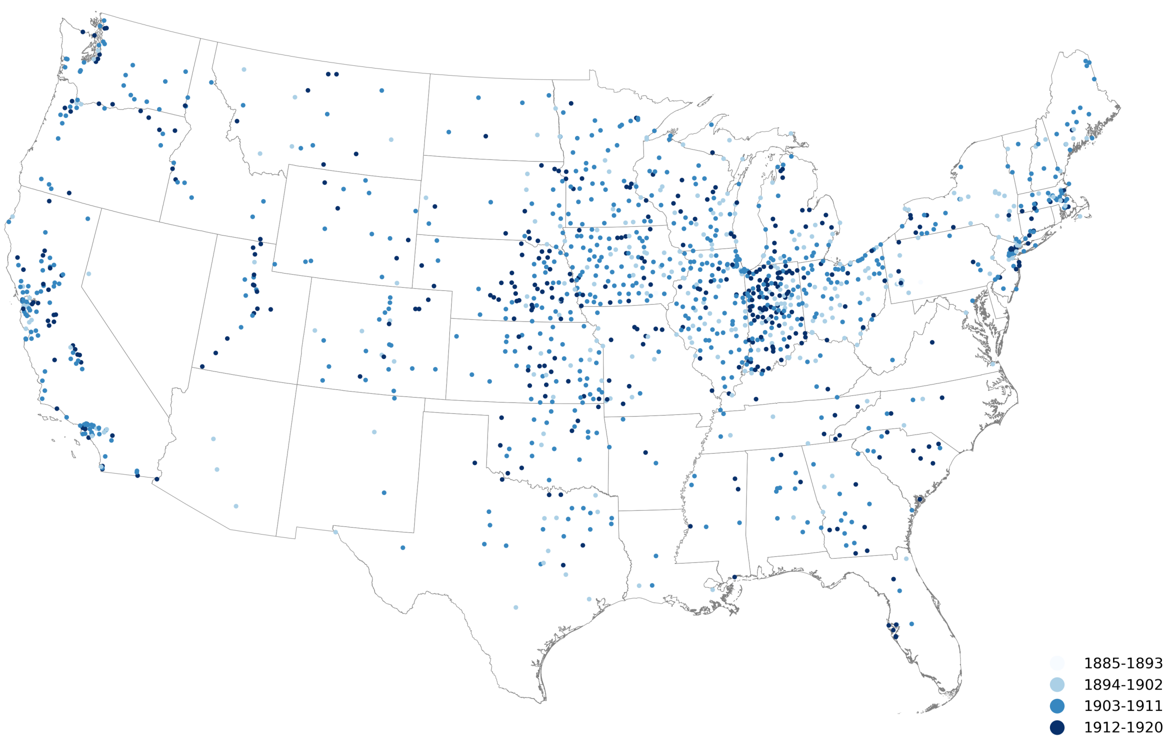
Notes: These figures show the cumulative distribution of accepted (Panel a) and rejected (Panel b) library cities by grant-year. The earliest grant year is used for cities that received multiple library grants. Source: Bobinski (1969) and Jones (1997).

Figure A2: **Map of all built and rejected Carnegie libraries**



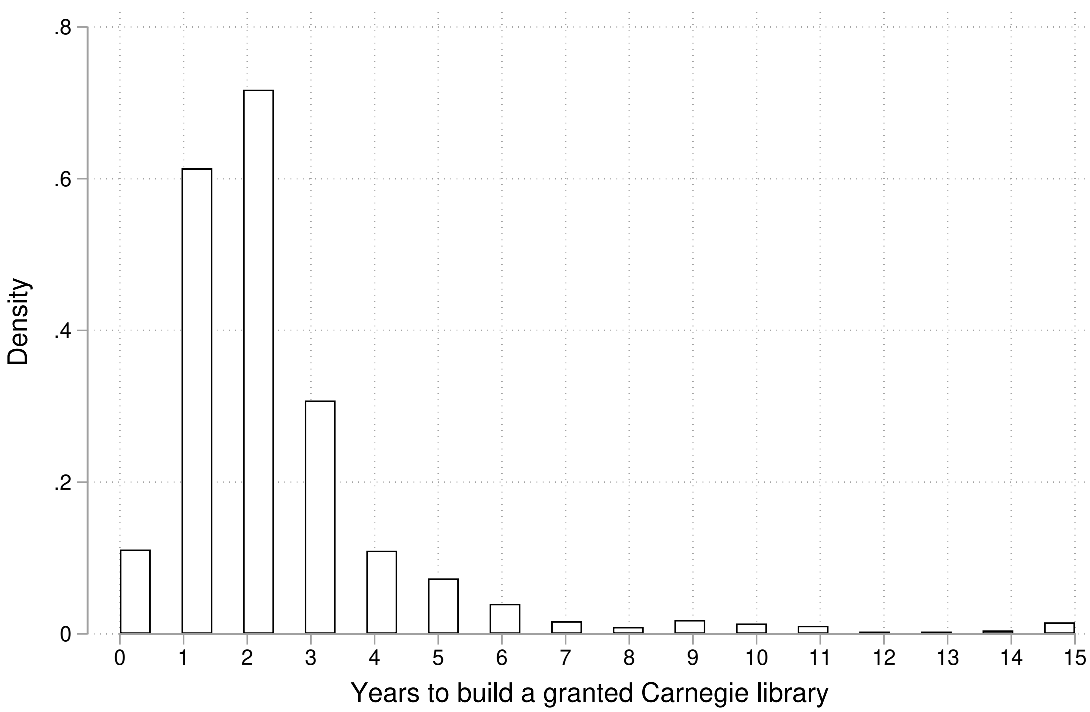
Notes: This figure shows the locations of all built and rejected Carnegie libraries. Blue (darker) dots correspond to libraries that were granted and built. Red dots with an "X" correspond to cities that rejected their Carnegie library grant. Sources: Bobinski (1969) and Jones (1997).

Figure A3: **Map of built Carnegie libraries over time**



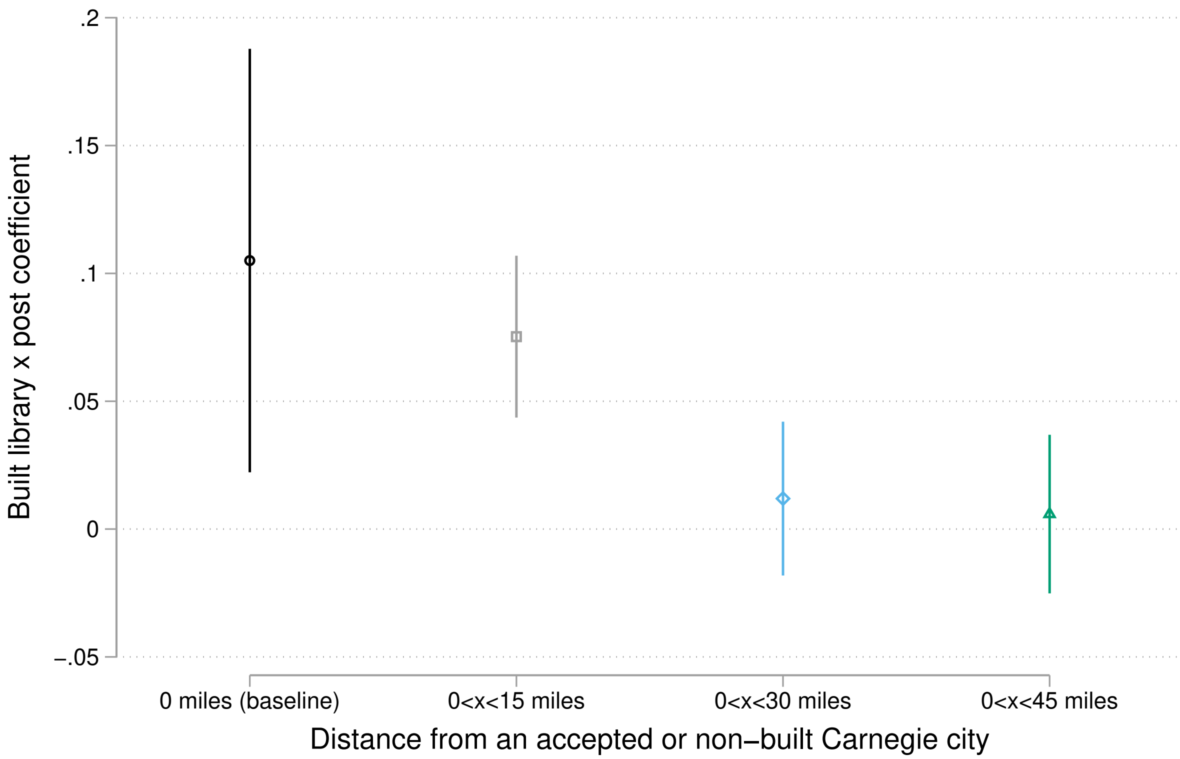
Notes: This figure shows the locations of all built Carnegie libraries. Grant years are divided into four groups. Darker dots correspond to places that received a grant later in the Carnegie program. Sources: Bobinski (1969) and Jones (1997).

Figure A4: **Distribution of time required to construct libraries after library grants**



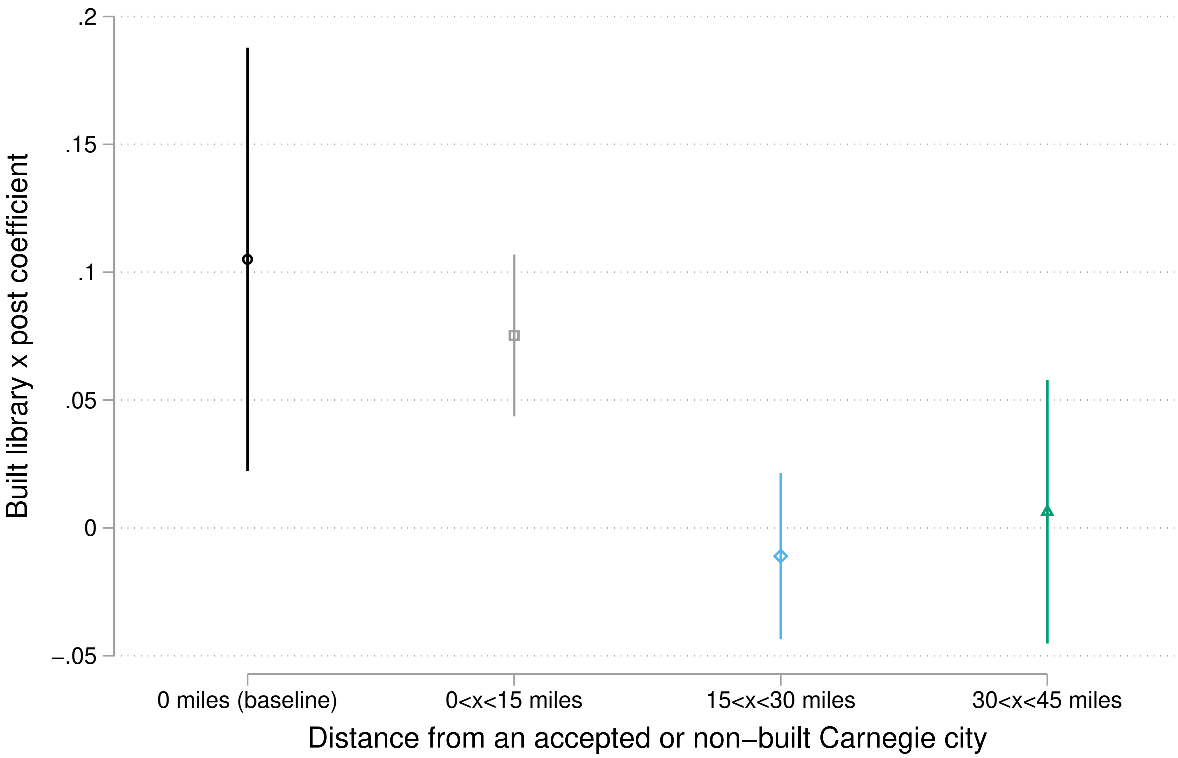
Notes: This figure shows the distribution of time (measured in years) between library grants and the libraries being opened to the public. The distribution is top-coded at 15 years. The average construction time was three years, but the most frequent construction time was between one and two years.

Figure A5: **Spillover effects of Carnegie libraries on patenting in nearby areas**



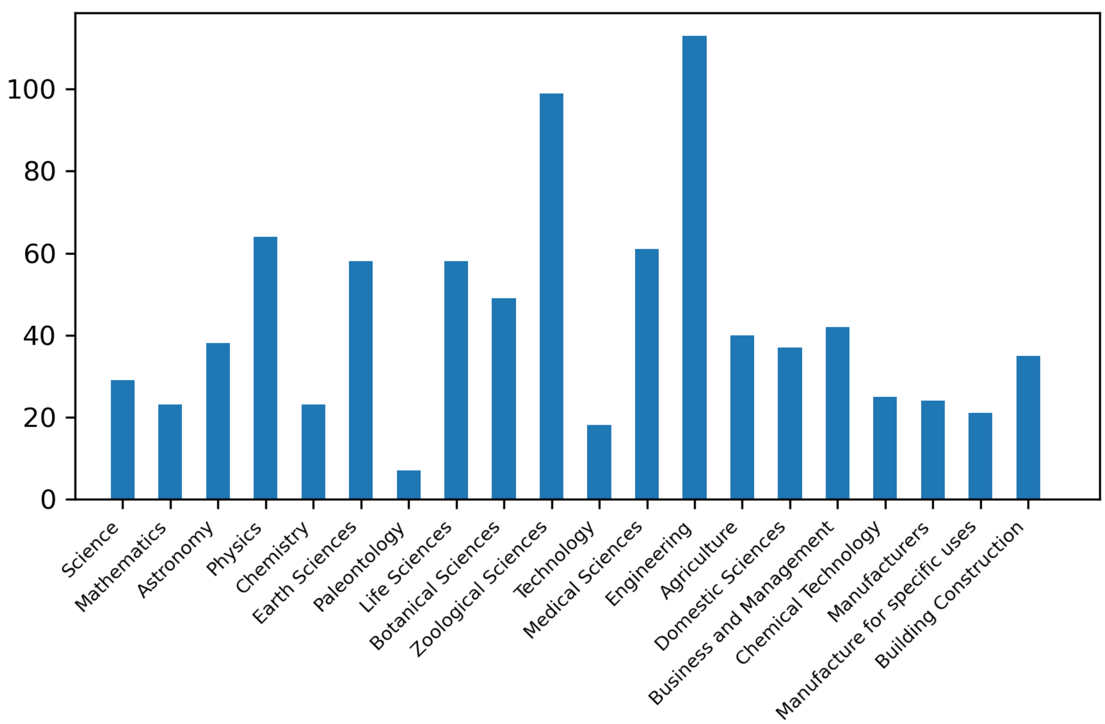
Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby places. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on city, state-year, and grant-year-by-year fixed effects. This estimate reflects 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. Standard errors are clustered at the grant-receiving-city level. 90% confidence intervals are shown.

Figure A6: **Spillover effects of Carnegie libraries on patenting in nearby places, non-intersecting distance bins**



Notes: This figure shows the marginal effects of receiving a Carnegie library on patenting in nearby places. The leftmost estimate is the baseline difference-in-differences estimate from regression Equation 1 conditional on on city, state-year, and grant-year-by-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. Standard errors are clustered at the grant-receiving-city level. 90% confidence intervals are shown.

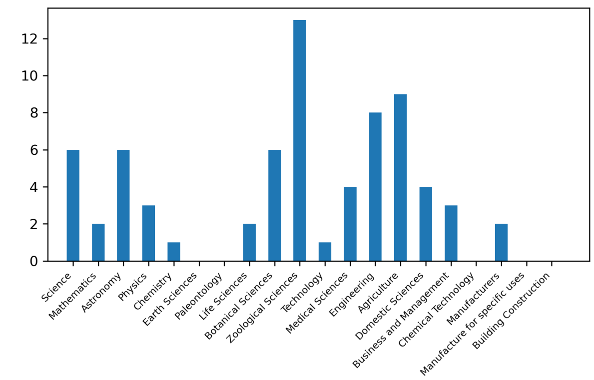
Figure A7: **Distribution of technical books in the 1904 ALA catalog**



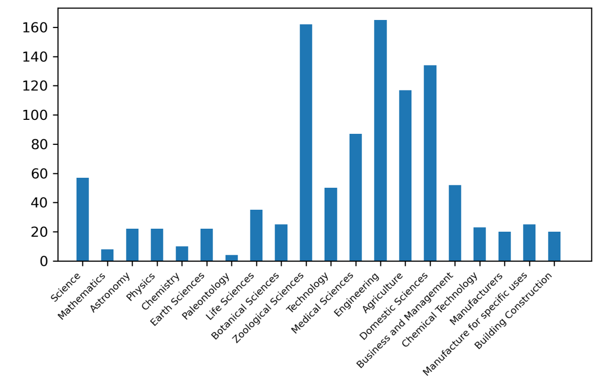
Notes: This figure shows the 2-digit Dewey Decimal Classification distribution of scientific books in the 1904 American Library Association (ALA) catalog.

Figure A8: **Distribution of technical books in the historical book catalog of 5 libraries**

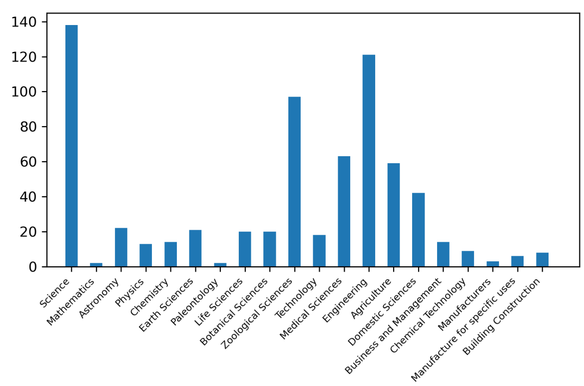
(a) Lexington, MI



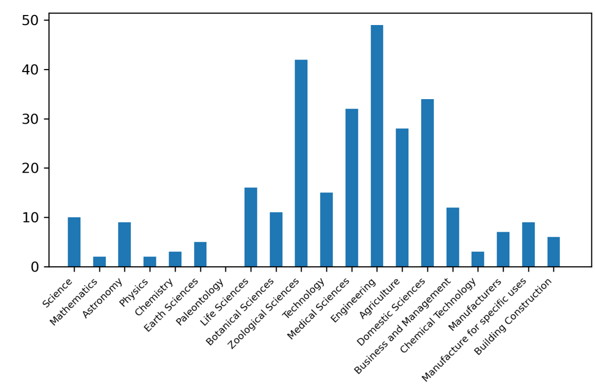
(b) Morris, IL



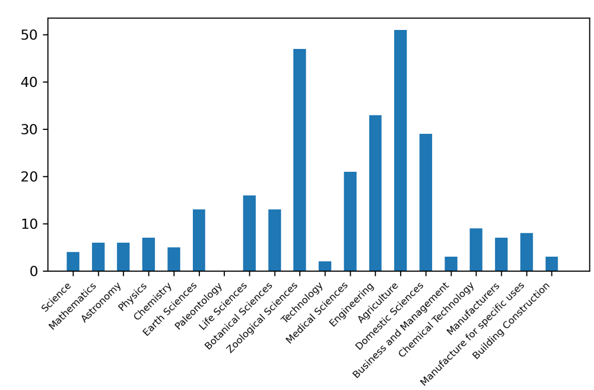
(c) Osage, IA



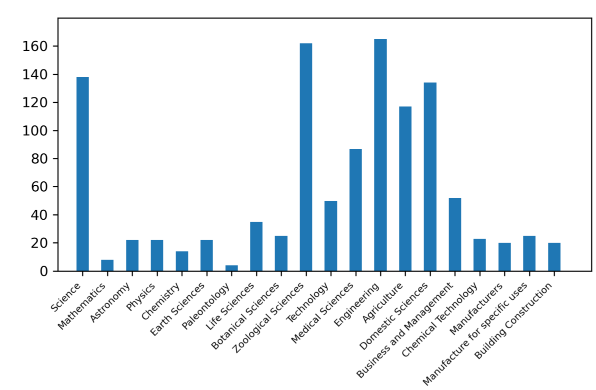
(d) Rhinelander, WI



(e) Sauk Center, MN



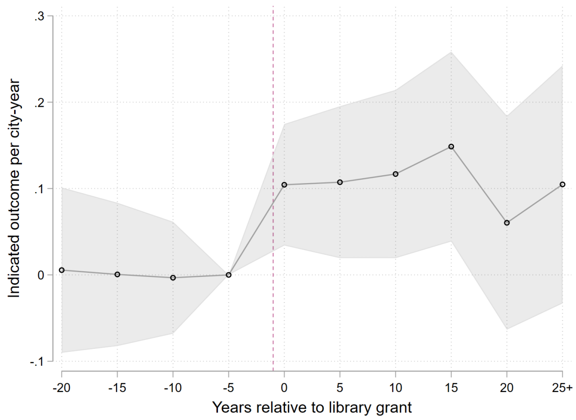
(f) Pooled Sample



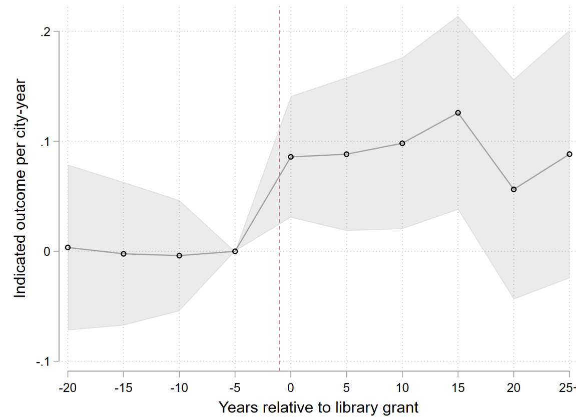
Notes: This figure shows the 2-digit Dewey Decimal Classification distribution for scientific books at five libraries obtained from the Main Street Public Library Dataset Wiegand (2011).

Figure A9: **Event study estimates of Carnegie libraries on patenting for alternative patent measures, models with city and state-year fixed effects**

(a) Inverse hyperbolic sine patents

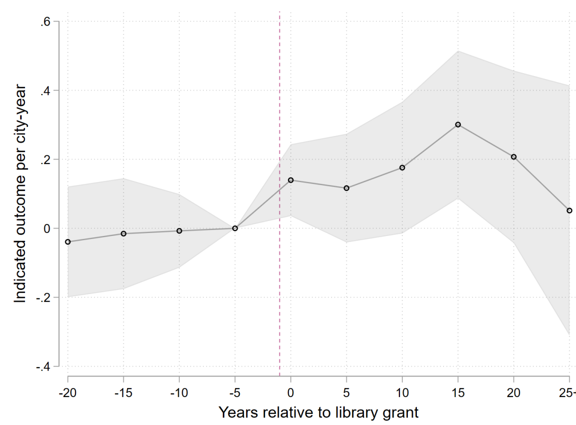


(b) Log patents + 1



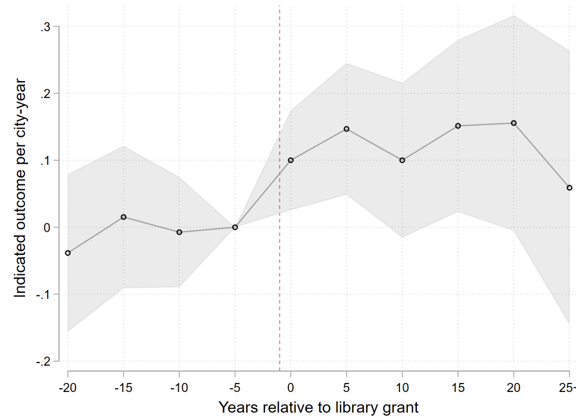
(c) Patent counts

(d) Patent counts (Poisson)



(e) Inverse hyperbolic sine, $pat > 0$

(f) Log patents, $pat > 0$

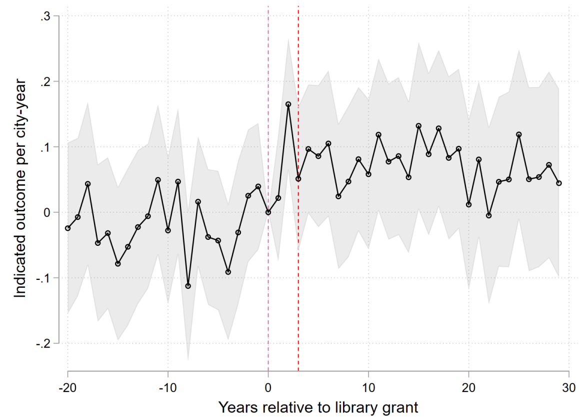


Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. Coefficients are binned in 5-year increments. All models include state-year and city fixed effects. Standard errors are clustered by city.

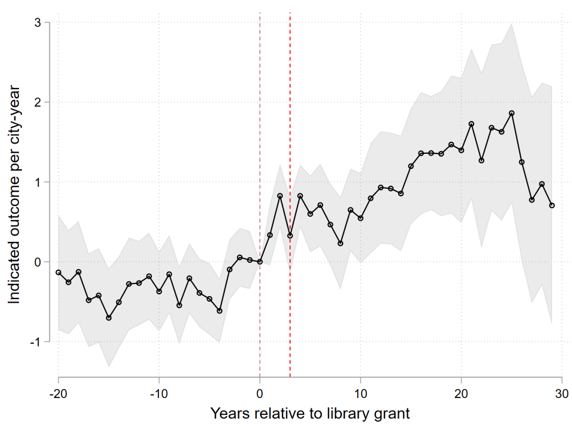
Figure A10: **Event study estimates of Carnegie libraries on patenting for alternative patent measures, models with city and state-year fixed effects (yearly)**

(a) Inverse hyperbolic sine patents

(b) Log patents + 1

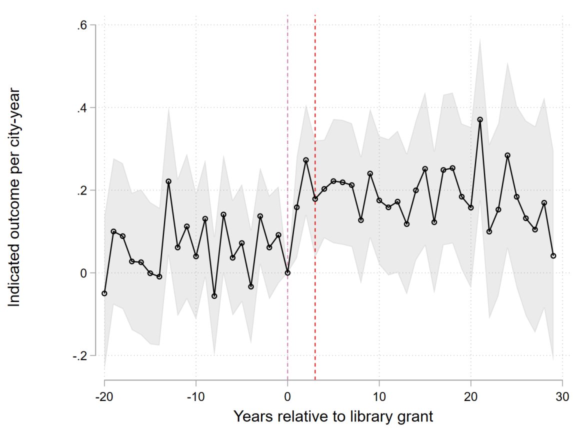


(c) Patent counts

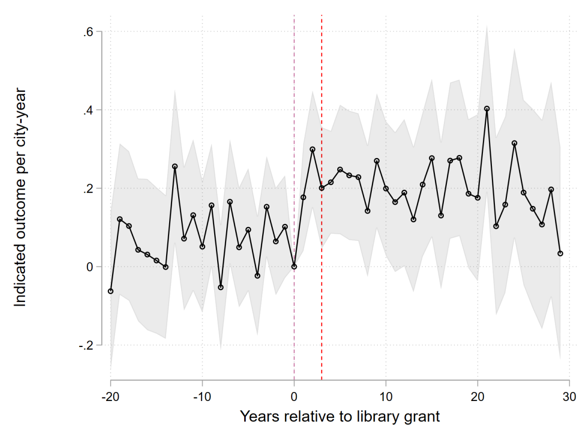


(d) Patent counts (Poisson)

(e) Inverse hyperbolic sine, $pat > 0$



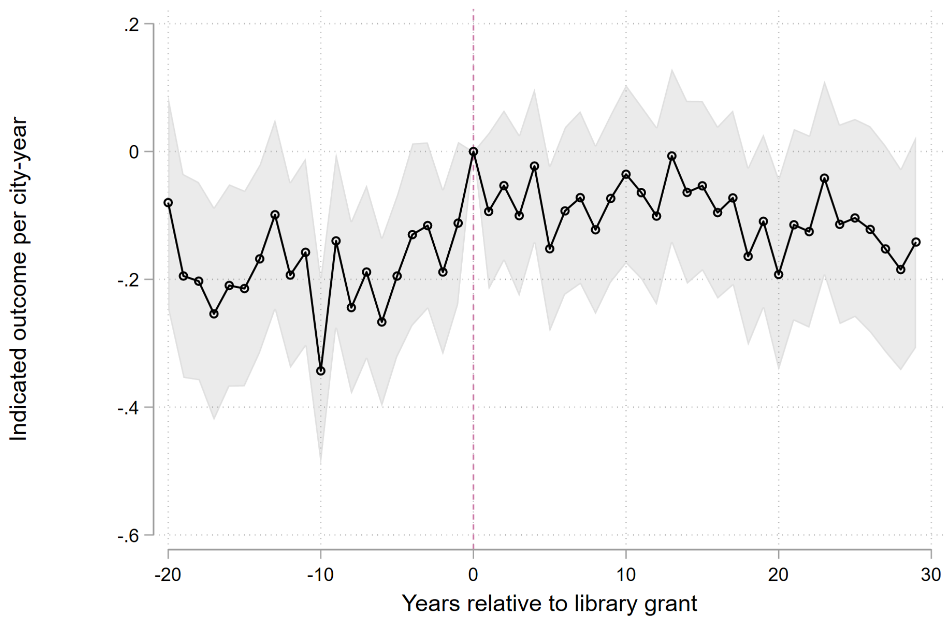
(f) Log patents, $pat > 0$



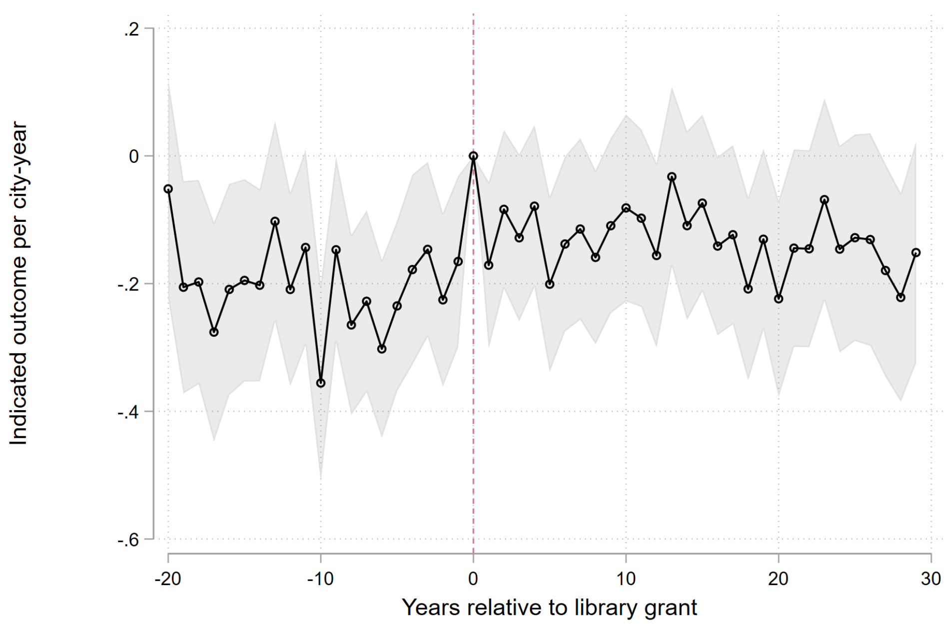
Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. All models include state-year, city, and grant-year-by-year fixed effects. Standard errors are clustered by city.

Figure A11: **Event study estimates of Carnegie libraries on city patenting, yearly estimates using opening dates**

(a) State-year fixed effects



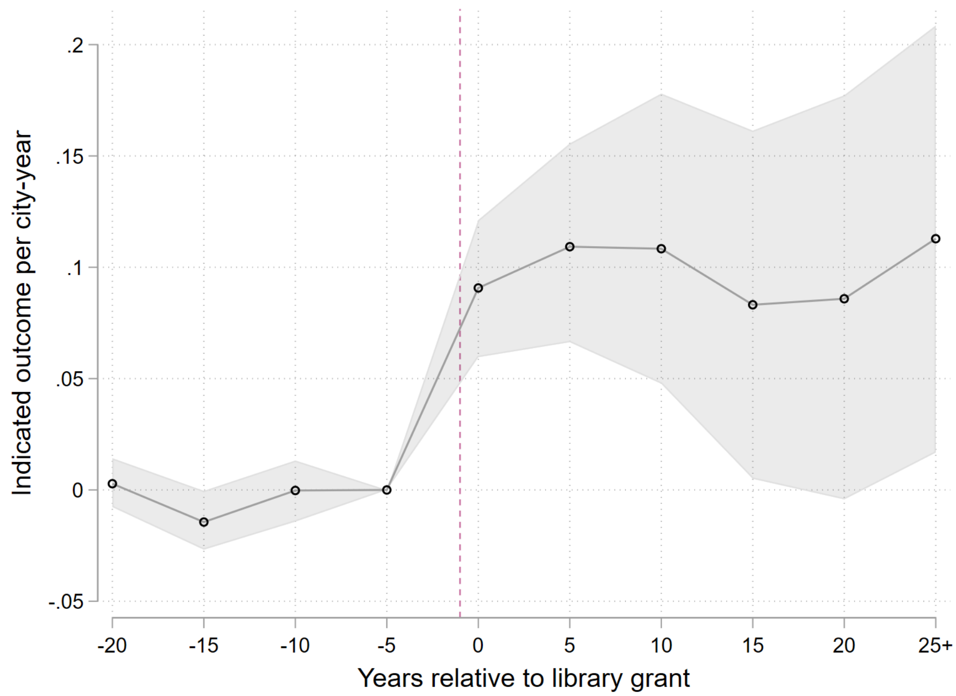
(b) State-year and grant-year-by-year fixed effects



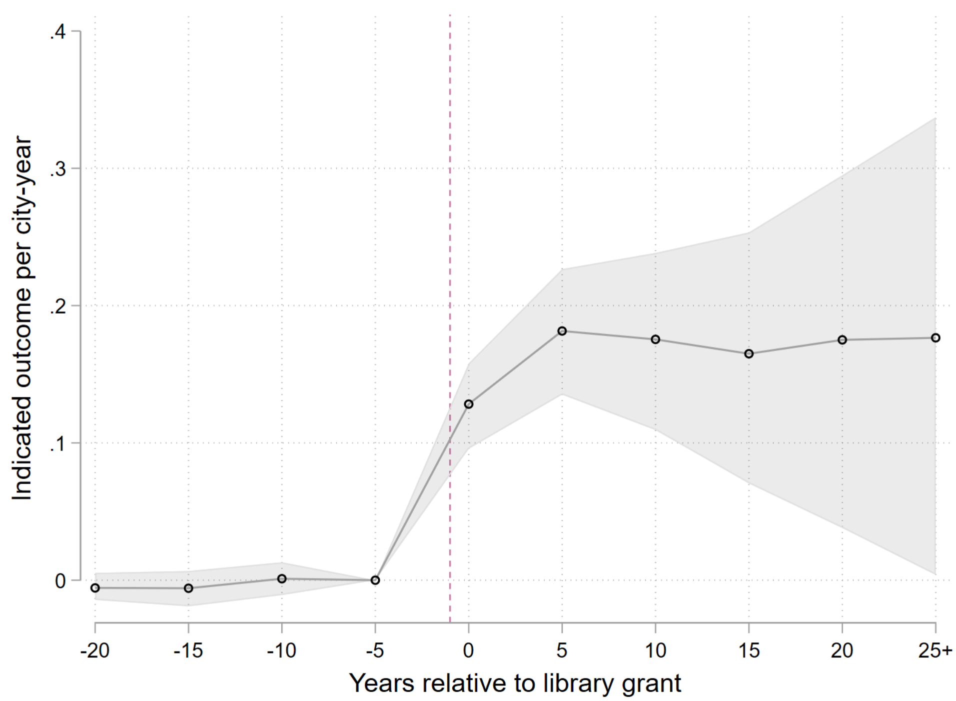
Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. We impute construction time for rejecting cities as the grant year plus two (the modal construction time). All models include state-year, city, and grant-year-by-year fixed effects. Standard errors are clustered by city.

Figure A12: **Event study estimates of Carnegie libraries on patenting, Gardner method**

(a) Inverse hyperbolic sine patents



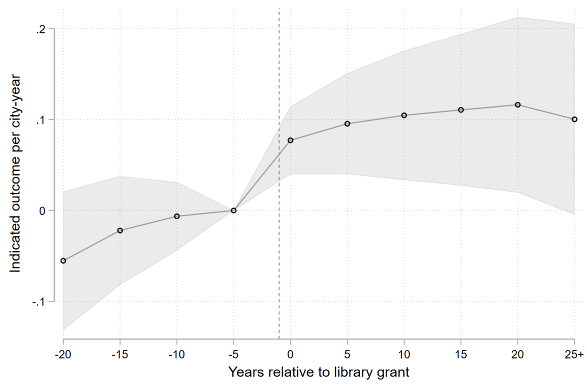
(b) Inverse hyperbolic sine patents, condition on $pat > 0$



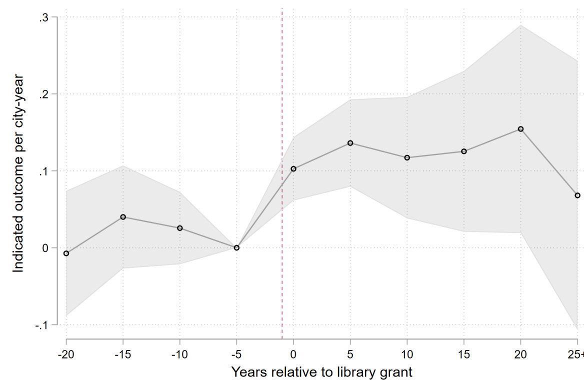
Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. Coefficients are binned in 5-year increments. Models are estimated using the two-stage Gardner method. Standard errors are calculated using the clustered Bayesian bootstrap with 1,000 repetitions, clustering at the city level.

Figure A13: **Event study estimates of Carnegie libraries on patenting for alternative patent measures, stacking approach**

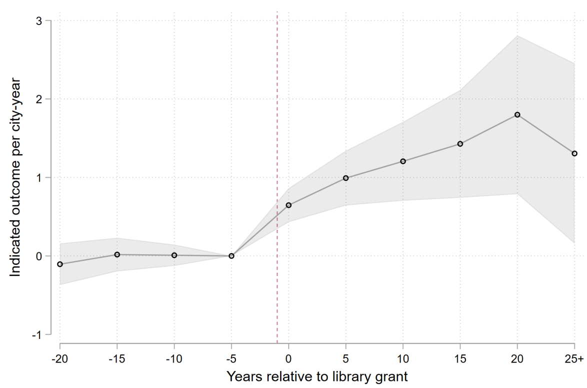
(a) IHS patents



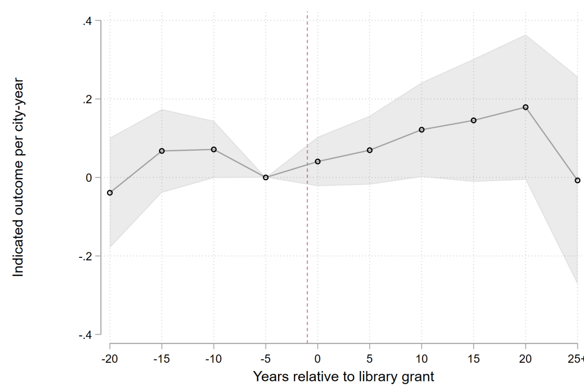
(b) IHS patents, conditional on any patent



(c) Patent counts



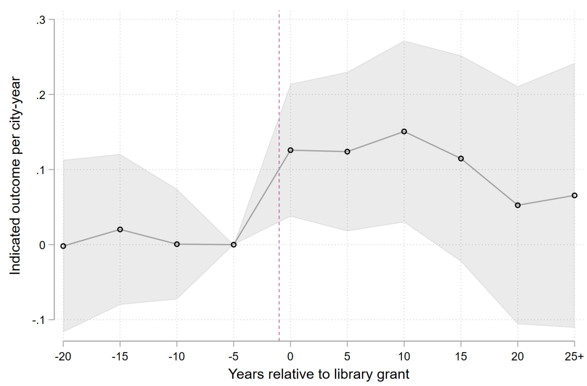
(d) Patent counts (Poisson)



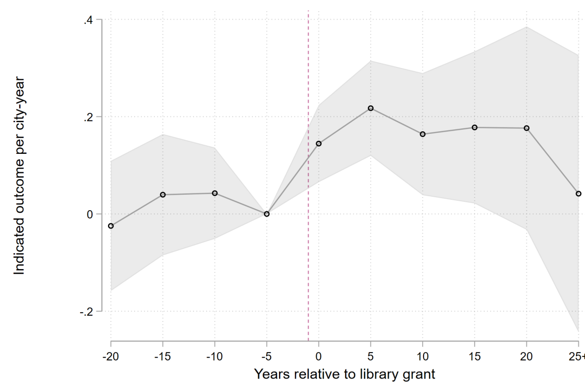
Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. Coefficients are binned in 5-year increments. Estimates are calculated using the stacking method, stacking each year treatment cohort with all control units. We include city and state-year fixed effects, interacting each set of fixed effects with "cohort" effects (i.e., grant-years). Standard errors are clustered at the city level.

Figure A14: **Event study estimates of Carnegie libraries on patenting for alternative patent measures, stacking approach using +1/-1 year window**

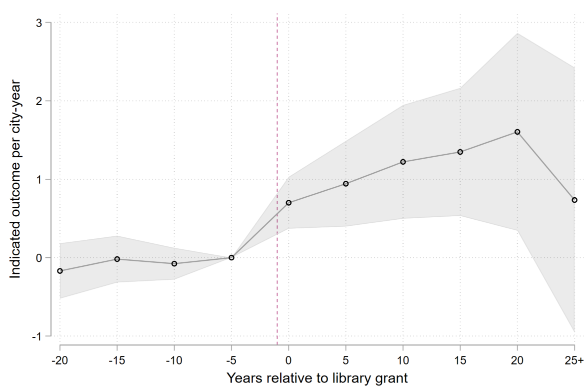
(a) IHS patents



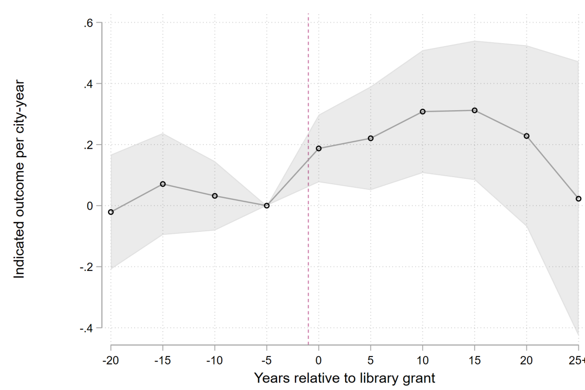
(b) IHS patents, conditional on any patent



(c) Patent counts

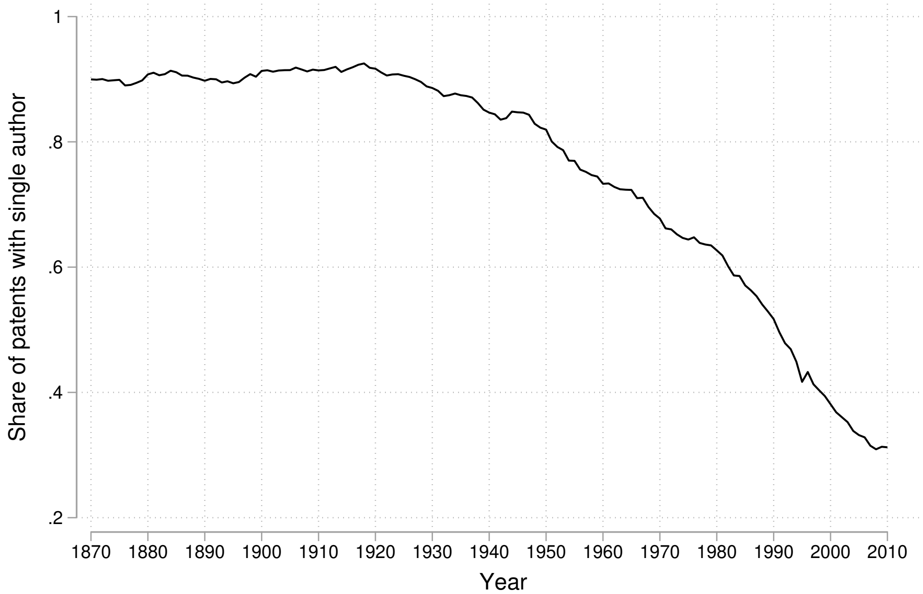


(d) Patent counts (Poisson)



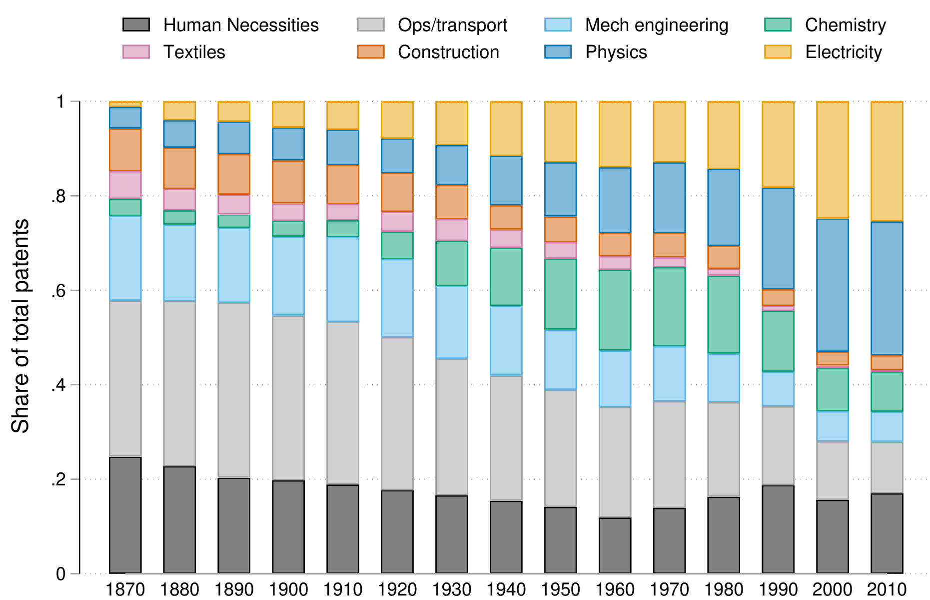
Notes: This figure shows the estimated effects of building a Carnegie library over time relative to grant years for different transformations of the outcome variable. Coefficients are binned in 5-year increments. Estimates are calculated using the stacking method, stacking year treatment cohort together with control units with grant years +1/-1 from a treatment cohort year. Estimates are calculated using the stacking method, stacking each year treatment cohort with all control units. We include city and state-year fixed effects, interacting each set of fixed effects with "cohort" effects (i.e., grant-years). Standard errors are clustered at the city level.

Figure A15: **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.

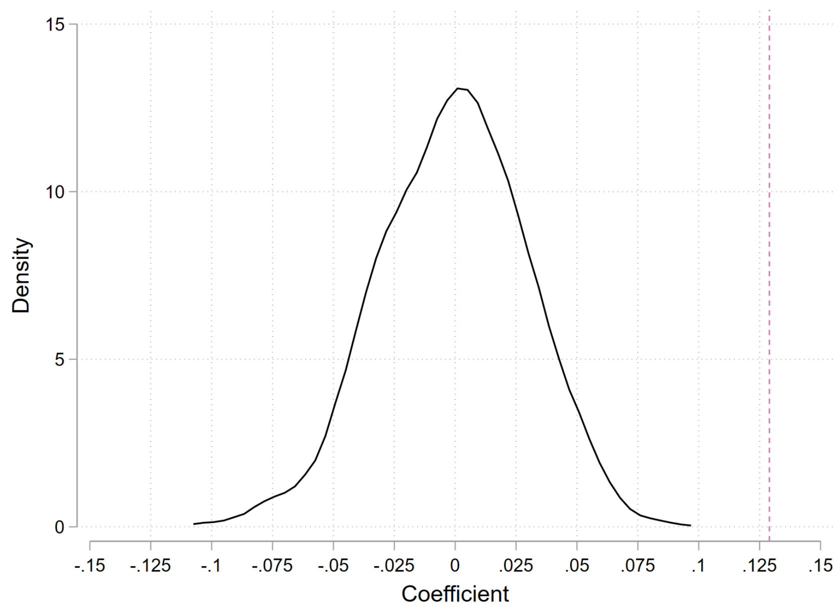
Figure A16: **Share of U.S. patents by Cooperative Patent Classification technology class by decade**



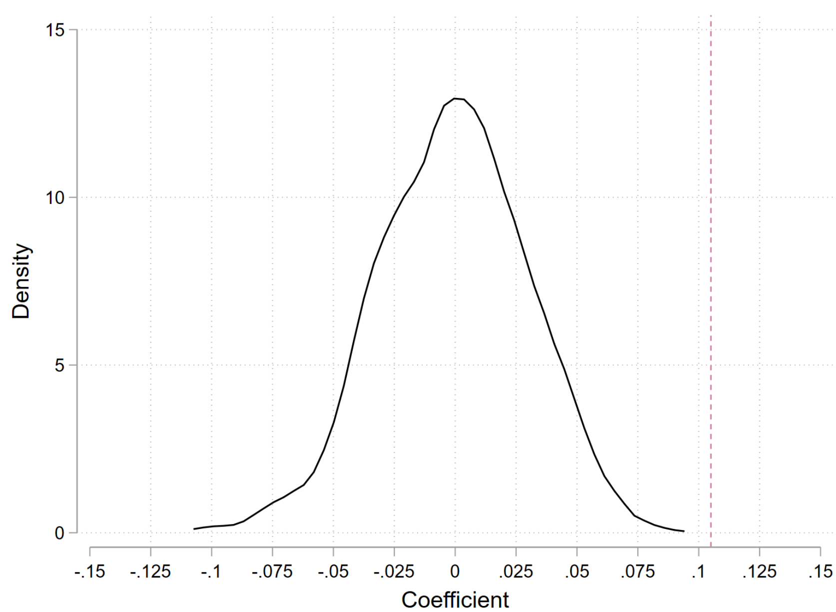
Notes: This figure shows the share of U.S. patents in each Cooperative Patent Classification technology class, by decade. Source: Comprehensive Universe of U.S. Patents (CUSP) data.

Figure A17: **Null treatment effect estimate distribution and actual treatment effect estimates (dashed line)**

(a) State-year fixed effects



(b) State-year and grant-year-by-year fixed effects



Notes: This figure shows a simulated null sampling distribution assuming no effect of libraries. We construct this distribution using a sample constructed to mimic the size of the baseline estimation sample and the ratio between treated and control cities. The methodology details are discussed in Appendix D. Vertical lines represent treatment effects obtained with the baseline regression model.

Table A1: Summary statistics

	Mean	Std. dev.	Min.	Max.
City-year variables (N = 48,698)				
Patent count	2.509	5.826	0.000	181.500
$ihs(patents)$	1.007	1.056	0.000	5.894
$ln(patents + 1)$	0.789	0.842	0.000	5.207
Female patents	0.094	0.331	0.000	14.125
Immigrant patents	0.208	0.534	0.000	16.946
Observed a patent that cited prior work	0.011	0.106	0.000	1.000
Patents that cite prior work	0.014	0.146	0.000	9.000
Observed a multi-inventor patent	0.165	0.371	0.000	1.000
Multi-inventor patents	0.244	0.851	0.000	37.000
Forward citations	2.680	12.216	0.000	694.833
Had a forward citation	0.297	0.457	0.000	1.000
Had a p90 breakthrough patent	0.081	0.272	0.000	1.000
Count first-time patents	1.248	2.266	0.000	64.000
1900 time-invariant variables (N = 1,221)				
Built Carnegie library	0.876	0.329	0.000	1.000
Share youth in-school	0.619	0.130	0.025	0.919
Share Black	0.057	0.139	0.000	1.000
Share female	0.493	0.037	0.204	0.570
Share miner	0.016	0.059	0.000	0.708
Occupation-industry earnings proxy	18.013	2.569	5.336	24.800
Population	4652.559	4574.500	83.000	24913.000
Average age	27.440	2.742	18.759	40.471
Share professional	0.081	0.033	0.000	0.254
Share managers	0.094	0.042	0.000	0.311
Share craftsmen	0.153	0.055	0.000	0.443
Share skilled operators	0.138	0.085	0.000	0.769
Strikers	566.559	3590.045	0.000	76949.000
Knights of Labor Assemblies	2.650	9.975	0.000	152.000
Had college	0.103	0.304	0.000	1.000
Share ag.	0.183	0.179	0.000	0.931
Share mining	0.019	0.066	0.000	0.739
Share construction	0.072	0.028	0.000	0.434
Share durable manufac.	0.051	0.056	0.000	0.540
Share non-durable manufac.	0.055	0.061	0.000	0.668
Share transport	0.087	0.066	0.000	0.566
Share retail/trade	0.118	0.042	0.000	0.314
Share finance/RE	0.027	0.025	0.000	0.207
Share services	0.176	0.057	0.016	0.400

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

Table A2: Effect of Carnegie library entry on local newspaper market and vote shares

Dependent variable	Lib. × post	Std. error	Cities
All papers			
Had paper	0.016	(0.024)	1,221
Papers per 1k	-0.007	(0.013)	1,207
Circulation per 1k	-48.275	(42.381)	1,207
Republican papers			
Had paper	0.018	(0.020)	1,221
Papers per 1k	0.007	(0.008)	1,207
Circulation per 1k	17.454	(19.233)	1,207
Democratic papers			
Had paper	-0.004	(0.020)	1,221
Papers per 1k	-0.005	(0.005)	1,207
Circulation per 1k	0.146	(11.432)	1,207
Pres. vote share			
Republican (county)	0.000	(0.005)	1,220

Notes: This table shows the impact of Carnegie libraries on potential "placebo" outcomes. Outcome variables are whether a given city-year observation had a daily newspaper, the number of newspapers per 1,000 residents, the circulation of all newspapers, and the county-level presidential election share held by Republicans. Outcome data come from Gentzkow et al. (2011). We include separate regressions for the newspaper outcomes for Republican and Democratic-supporting newspapers.

Table A3: Robustness of difference-in-differences results to alternative specifications

Dependent variable and estimation strategy	Lib. $\times$ post	Std. error	Cities
Baseline $ihs(pat)$ model with s-y, gy-y fixed effects	0.105	(0.050)	1,221
Condition on time-varying log population	0.088	(0.048)	1,208
Patents per 1k people outcome	0.091	(0.043)	1,208
Condition on GY-Year $\times$ covariates	0.132	(0.068)	1,221
Condition on GY-Year $\times$ covariates & Had college	0.121	(0.069)	1,221

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 using alternative specifications and sets of controls. All models condition on state-year, city, and grant-year-by-year fixed effects. The covariates are the 1900 share of women, share of youth in school, share of Black people, population, share of miners, average imputed earnings, average age, share of workers in 1-digit occupation categories, share of workers in 1-digit industry categories, and the number of Knights of Labor assemblies and strikers.

Table A4: Effect of Carnegie libraries on measures of patent quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Had forward citation								
Built library × post	0.045	0.045	0.051	0.051	0.051	0.071	0.066	0.051
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
Count citations								
Built library × post	0.970	0.964	1.057	1.047	1.044	2.220	2.243	1.389
	(0.651)	(0.651)	(0.658)	(0.659)	(0.660)	(0.750)	(0.798)	(0.790)
Had top innovative pat.								
Built library × post	-0.005	-0.005	-0.004	-0.004	-0.004	0.014	0.012	0.002
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean had forward citation	0.297	0.297	0.297	0.297	0.297	0.297	0.297	0.297
Mean forward citations	2.680	2.680	2.680	2.680	2.680	2.680	2.680	2.680
Mean had top innovative pat.	0.081	0.081	0.081	0.081	0.081	0.081	0.081	0.081
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

Notes: This table shows the estimated impact of Carnegie libraries on measures of patent quality relative to a set of cities that were approved to build a library but did not do so. We report the estimates obtained 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 is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use 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). Each observation is at the city-year level. Standard errors are shown in parentheses and clustered by city.

Table A5: Effect of Carnegie libraries on women and immigrant patenting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Women, count								
Built library × post	0.021	0.021	0.021	0.020	0.020	0.037	0.037	0.022
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.016)	(0.016)
Women, share								
Built library × post	-0.002	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Immigrants, count								
Built library × post	0.083	0.082	0.080	0.080	0.080	0.124	0.125	0.087
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.030)	(0.033)	(0.032)
Immigrants, share								
Built library × post	0.000	0.000	-0.000	-0.000	-0.002	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.003)	(0.003)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean women patents	0.094	0.094	0.094	0.094	0.094	0.094	0.094	0.095
Mean immigrant patents	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

Notes: This table shows the estimated impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. We report the estimates obtained using a sample window of 20 years before and after Carnegie grants. Built library indicates cities that built a Carnegie library. Post is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use the earliest grant year. The outcome variable is the count and share of patents for women and immigrants, each identified using the name-based procedure described in Appendix C. Each observation is at the city-year level. Standard errors are shown in parentheses and clustered by city.

Table A6: Effect of Carnegie libraries on patenting, extensive margin ($Pat > 0$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.038	0.038	0.036	0.036	0.036	0.026	0.017	0.025
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.022)	(0.022)	(0.022)
Pre-1929 sample								
Built library × post	0.049	0.046	0.044	0.043	0.037	0.034	0.018	0.027
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.022)	(0.021)	(0.022)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean $Pr(Pat > 0)$	0.593	0.593	0.593	0.593	0.593	0.593	0.593	0.593
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

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. Built library indicates cities that built a Carnegie library. Post is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use the earliest grant year. The outcome variable is $Pr(Patents > 0)$. Each observation is at the city-year level. Standard errors are shown in parentheses and clustered by city.

Table A7: 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	0.287	0.283	0.279	0.276	0.276	0.361	0.336	0.195
	(0.109)	(0.109)	(0.108)	(0.108)	(0.109)	(0.111)	(0.112)	(0.117)
Share first-time								
Built library × post	-0.006	-0.008	-0.005	-0.005	-0.012	-0.008	-0.014	-0.010
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.020)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean first-time count	1.248	1.248	1.248	1.248	1.248	1.248	1.248	1.248
Mean share first-time	0.629	0.629	0.629	0.629	0.629	0.629	0.629	0.629
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

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. Built library indicates cities that built a Carnegie library. Post is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use the earliest grant year. The outcome variable is indicated in each panel. Each observation is at the city-year level. Standard errors are shown in parentheses and clustered by city.

Table A8: Heterogeneity in library difference-in-differences estimates across city characteristics

Coefficient	Estimate	Standard error
Built library $\times$ post $\times$ had college	0.093	(0.187)
Built library $\times$ post $\times$ share in-school top half	0.155	(0.094)
Built library $\times$ post $\times$ imputed income top half	0.073	(0.095)
Built library $\times$ post $\times$ share craftsmen top half	-0.054	(0.096)
Built library $\times$ post $\times$ share Black top half	-0.000	(0.094)
Built library $\times$ post $\times$ population top half	0.178	(0.100)

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 $ihs(patents)$ as the outcome variable. All models include city, state-year, and grant-year-by-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. Regression models 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. Each observation is at the city-year level. Standard errors are clustered by city.

Table A9: Robustness of difference-in-differences patent results to alternative samples and post period lengths

Patent class	S-Y FE		S-Y and GY-Y FE		Cities
	Lib. × post	Std. err.	Lib. × post	Std. err.	
20-year post-period					
Baseline model	0.129	(0.050)	0.105	(0.050)	1,221
Excluding observations					
Exclude Southern states	0.163	(0.054)	0.143	(0.056)	1,060
Exclude New York and Pennsylvania	0.119	(0.051)	0.090	(0.052)	1,152
15 year pre-period	0.125	(0.048)	0.106	(0.049)	1,221
10 year pre-period	0.125	(0.048)	0.112	(0.049)	1,221
Exclude "finance" motivated rejectors	0.134	(0.062)	0.110	(0.062)	1,173
Exclude cities larger than 15,000 people	0.121	(0.050)	0.104	(0.051)	1,158
Exclude cities larger than 5,000 people	0.108	(0.052)	0.105	(0.053)	866
Adding observations					
Include pre-1899 grants	0.123	(0.050)	0.102	(0.050)	1,225
Include large population cities	0.157	(0.050)	0.126	(0.050)	1,327
Include library philanthropist cites	0.112	(0.048)	0.085	(0.049)	1,242
Include cities missing 1900 covariates	0.084	(0.047)	0.069	(0.048)	1,450
Include all cities	0.102	(0.044)	0.081	(0.045)	1,608
10-year post-period					
Baseline model	0.115	(0.045)	0.095	(0.046)	1,221
Excluding observations					
Exclude Southern states	0.138	(0.050)	0.122	(0.052)	1,060
Exclude New York and Pennsylvania	0.118	(0.046)	0.093	(0.048)	1,152
15 year pre-period	0.111	(0.042)	0.094	(0.044)	1,221
10 year pre-period	0.115	(0.042)	0.101	(0.043)	1,221
Exclude "finance" motivated rejectors	0.136	(0.054)	0.114	(0.055)	1,173
Exclude cities larger than 15,000 people	0.110	(0.045)	0.093	(0.046)	1,158
Exclude cities larger than 5,000 people	0.106	(0.046)	0.100	(0.047)	866
Adding observations					
Include pre-1899 grants	0.110	(0.045)	0.092	(0.046)	1,225
Include large population cities	0.136	(0.045)	0.110	(0.046)	1,327
Include library philanthropist cites	0.100	(0.043)	0.076	(0.044)	1,242
Include cities missing 1900 covariates	0.084	(0.041)	0.072	(0.042)	1,450
Include all cities	0.096	(0.039)	0.079	(0.040)	1,608

Notes: This table shows robustness 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 in 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. Each observation is at the city-year level. Standard errors are clustered by city.

Table A10: Robustness of difference-in-differences results to alternative patent transformations and estimation strategies

Patent class	S-Y FE		S-Y and GY-Y FE	
	Lib. $\times$ post	Std. err.	Lib. $\times$ post	Std. err.
$ihs(patents)$	0.129	(0.050)	0.105	(0.050)
$ln(patents + 1)$	0.109	(0.040)	0.088	(0.041)
Patent counts	1.284	(0.364)	0.932	(0.356)
Poisson counts	0.221	(0.105)	0.226	(0.107)

Notes: This table shows robustness results from our baseline model using different transformations of the outcome variables and estimation strategies. All models include city fixed effects and the additional fixed effects indicated in the relevant column. Each observation is at the city-year level. All standard errors are clustered by city.

Table A11: Effect of Carnegie libraries on patenting, log patents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.098	0.096	0.092	0.091	0.091	0.119	0.109	0.088
	(0.039)	(0.039)	(0.038)	(0.038)	(0.038)	(0.042)	(0.040)	(0.041)
Pre-1929 sample								
Built library × post	0.124	0.117	0.111	0.108	0.095	0.142	0.114	0.093
	(0.040)	(0.040)	(0.039)	(0.039)	(0.038)	(0.043)	(0.040)	(0.041)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean $ln(patents + 1)$	0.789	0.789	0.789	0.789	0.789	0.789	0.789	0.789
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

Notes: This table shows the estimated impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. We report the estimates obtained 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 is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use 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 A12: Effect of Carnegie libraries on patenting, patent counts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample								
Built library × post	0.791	0.783	0.762	0.755	0.755	1.261	1.284	0.932
	(0.274)	(0.274)	(0.269)	(0.269)	(0.270)	(0.344)	(0.364)	(0.356)
Pre-1929 sample								
Built library × post	0.906	0.867	0.842	0.824	0.752	1.368	1.298	0.929
	(0.293)	(0.291)	(0.286)	(0.285)	(0.276)	(0.362)	(0.371)	(0.362)
State FE	✓	✓	✓	✓		✓	✓	✓
Grant year FE		✓	✓	✓				✓
Cal. year FE			✓	✓	✓	✓	✓	✓
1900 covariates				✓				
City FE					✓		✓	✓
State-year FE						✓	✓	✓
Grant-Cal. year FE								✓
Mean patents	2.509	2.509	2.509	2.509	2.509	2.509	2.509	2.509
Observations	48,698	48,698	48,698	48,698	48,698	48,698	48,698	48,698
Cities	1,221	1,221	1,221	1,221	1,221	1,221	1,221	1,221

Notes: This table shows the estimated impact of Carnegie libraries on patenting relative to a set of cities that were approved to build a library but did not do so. We report the estimates obtained 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 is equal to 1 in the years after cities received Carnegie library grants. If a city received multiple grants, we use the earliest grant year. The outcome variable is patent counts. Each observation is a city-year. Standard errors are shown in parentheses and clustered by city.

Table A13: Effect of Carnegie libraries on patenting (aggregated model)

	$ihs(pat)$		$ln(pat+1)$		Poisson (count)	
Built library × post	0.243	0.200	0.220	0.178	0.274	0.218
	(0.088)	(0.088)	(0.080)	(0.079)	(0.116)	(0.117)
City FE	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓
State FE - Post FE		✓		✓		✓
Grant-year FE - Post FE		✓		✓		✓
Mean patent measure	3.721	3.721	3.121	3.121	50.028	50.028
Observations	2,442	2,442	2,442	2,442	2,442	2,442
Cities	1,221	1,221	1,221	1,221	1,221	1,221

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. Each observation is a city-year. Standard errors are clustered by city.

B Historical background appendix

In this appendix we provide additional information on the historical background and context that informs this project. This appendix supplements Section 2, which describes the major details of Carnegie's program.

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.

B.1 Public libraries before Carnegie

Public libraries in the United States are a relatively recent civic institution. In 1833, the small city of Petersborough, N.H., established the first public library in the United States, open to all citizens and supported by city 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. Public libraries competed with subscription libraries. Often these subscription libraries carried primarily fiction texts, but some private libraries were also oriented toward engineering topics (e.g., the Mechanics Institutes).

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 advocated 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 a city. Over half of these new libraries were constructed with funds from a single donor: Andrew Carnegie.

B.2 Carnegie's grant process

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.[22] 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.[23]

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

[23]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

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.[24]

B.3 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 A15 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 steadily declined, reaching about 35 percent in 2010. Figure A16 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

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. (e.g., Bayliss, California).

[24]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. This net effect conflates a quantity and quality dimension.

(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 reference and information materials that libraries held.

C 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.

C.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 "Carnegie Libraries of California."[25] Since the

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

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 Collections 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 opened 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 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.[26]

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

C.2 Patent data

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.[27] 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.

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

[27] Berkes (2018) provides a full description of these data and how it was compiled. Andrews (2021) 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.

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.

We also assign predicted probabilities that authors on each patent are female or an immigrant. From the patent data, we observe the exact name of inventors. To assign gender and immigrant 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. We assign this proportion to each corresponding patent name. 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. We perform a similar exercise for immigrants, but we use the last name since many immigrants change first names upon arrival.

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 patent texts and identify those that mention these keywords. We manually review the matches and extend the set of keywords based on our inspection. 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 not 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);[28] *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."[29]

Covariates

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

[28]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.

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

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.[30]

For robustness, we use two county-level proxies for union activity measured before 1900. First, 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).[31] 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.[32] 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

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

[31]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.

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

universities.[33]

D Power analysis appendix

In this section, we discuss a number of analyses to discuss the *ex-post* power of our baseline statistical approaches. Given our empirical strategy, formula-based power approaches can be difficult, so we focus on two related, simulation-based exercises.

First, we have undertaken a placebo analysis based on the assumption of a null effect. We follow the following procedure:

1. Start with our patenting data and drop actual treatment and rejecting Carnegie cities.
2. Limit the remaining sample to places bigger than 1,000 people in 1900 that satisfy our baseline sample requirements (e.g., we can identify them in the census data).
3. Assign cities to treatment or control group at random to match the number of places and proportion of treatment vs. control places in our baseline sample
4. Assign grant dates to these cities between 1899 and 1919.
5. Assign post-period treatment effects of ”0” with a standard deviation around that assignment of 0.02 IHS points
6. Estimate our baseline models, saving the coefficients.
7. Repeat steps 1-6 for 1000 iterations

[33] 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.

Figure A17 shows the kernel distribution of treatment effects under these "null" scenarios and compares them to our baseline treatment effect estimates (vertical lines) from our actual sample. Figure A17 shows that in all cases, our treatment effect is at the far end (or outside of) the distribution of placebo estimates, suggesting that it is unlikely that our results emerge due to chance or an outlier that could drive the results in an underpowered analysis.

Next, we perform a more traditional power analysis. We follow a similar set of steps:

1. Start with our patenting data and drop actual treatment and rejecting Carnegie cities.
2. Limit the remaining sample to places bigger than 1,000 people in 1900 that satisfy our baseline sample requirements (e.g., merge to the census, are not too large).
3. Assign treatment and control cities at random to match the number of places and proportion of treatment vs. control places in our baseline sample
4. Assign grant dates to these cities between 1899 and 1919.
5. Assign post-period treatment effects drawing from a normal distribution with mean TE and standard deviation 0.02, where TE is a number corresponding to our baseline state-year estimate or our state-year and grant-year-by-year fixed effects estimates.
6. Estimate our baseline models, saving the resulting treatment effect t-statistics.
7. Repeat steps 1-6 for 500 iterations, and calculate the proportion of iterations where the t-statistics allow us to reject the null hypothesis (no treatment effect) at a 5% and 10% significance level. This proportion is what we report as "power."

Following this calculation, we find that our treatment in a model with state-year fixed effects can be detected with power 98% when using a 10% significance level or 97% when using a 5%

significance level. Our state-year and grant-year-by-year model's baseline effect size can be detected with power 93% using a 10% significance level or 89% power using a 5% significance level. These figures are well above the commonly cited power threshold of 80% used as a baseline for evaluating both *ex ante* and *ex post* power (e.g., Black et al., 2022).

To see how sample size plays a role in our analysis, we try limiting the number of control cities in the simulation to 100, we lose some power. However, even such a large drop in our control sample size leaves us reasonably well-powered to detect the effects we measure in our baseline analyses. In particular, our state-year effect can be detected with power 96% when using a 10% significance level or 90% using a 5% significance level. Our state-year and grant-year-by-year model's baseline effect size can be detected with power 84% using a 10% significance level or 76% power using a 5% significance level.