Andrews, Michael J. 2023. "How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of US Colleges." American Economic Journal: Economic Policy, 15 (2): 1–41. DOI: 10.1257/pol.20200320

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges[*]

Michael Andrews[†]

March 28, 2020

Abstract

I use data on site selection decisions for a subset of U.S. colleges to identify "runner-up" locations that were strongly considered to become the sites of new colleges but were ultimately not chosen for as-good-as-random reasons. When using runner-up counties as counterfactuals, establishing a college causes 48% more patents per year. Linking patents to novel college yearbook data reveal that only 5% of patents in a college's county came from alumni or faculty of that college. I find no difference in patenting between establishing colleges and establishing other types of institutions, nor between colleges with different focuses on technical fields.

---

[*] I am grateful to the Kauffman Foundation, NSF Doctoral Dissertation Grant #1661421, the Balzan Foundation, and the Northwestern Center for Economic History for financial support. I am also indebted to the librarians in the Interlibrary Loan Office at the University of Iowa and Northwestern University, who have responded to a seemingly endless series of requests for old books about even older universities. I would also like to thank Enrico Berkes, Chelsea Crain, Claudia Goldin, Shawn Kantor, John Parman, Elisabeth Perlman, Heyu Xiong, and Yiling Zhao for sharing data and/or code. In addition to those mentioned above, this paper benefited from conversations with Ufuk Akcigit, Wes Cohen, Mercedes Delgado, Arin Dube, Joe Ferrie, Roberto Fontana, Michael Frakes, Walker Hanlon, Ben Jones, Shawn Kantor, Megan MacGarvie, Joel Mokyr, Laura Owens, Sarada, Scott Stern, Alex Whalley, and Nicolas Ziebarth, as well as from seminar and conference participants at the University of Iowa, Kauffman Foundation, Northwestern University, NBER Innovation and DAE Summer Institutes, University of Wisconsin, Searle Center Conference on Innovation Economics, Auburn University, Economic History Association Annual Meetings, NBER Productivity lunch, Illinois Economic Association, Chicago Fed, Duke University, Utah State University, Bureau of Economic Analysis, Cliometric Society Annual Conference, LSE Innovation conference, Ohio State University, Rensselaer Poyltechnic Institute, Roundtable for Engineering and Entrepreneurship Research, Conference on Urban and Regional Economics, Duke University Conference on Empirical Patent Law, and the USDA Economic Research Service. All errors are my own. The most recent version of this paper is available at https://www.m-andrews.com/research.

[†] National Bureau of Economic Research. 1050 Massachusetts Ave., Cambridge, MA 02138. *Email*: mandrews@nber.org.

How do institutions of higher education affect local invention?[1] The literature proposes many possible channels, including both direct and indirect channels. In direct channels, activities that are unique to colleges, namely students obtaining human capital and faculty conducting research, directly push forward the frontier of practical knowledge.[2] In indirect channels, colleges instead affect the local innovation ecosystem in more subtle ways, with individuals who are not directly affiliated with a college benefiting from the presence of the institution to produce inventions. These indirect channels can take many forms; for example, the college may drive migration that in turn generates agglomeration externalities (Carlino and Kerr, 2015; Ciccone and Hall, 1996; Glaeser and Gottlieb, 2009). Attempts to estimate the causal effect of colleges on local invention (Aghion, Boustan, Hoxby, and Vandenbussche, 2009; Furman and MacGarvie, 2007; Hausman, 2017; Kantor and Whalley, 2014; Kim, 2020) typically shed little light on the channels through which colleges operate. In this study, I exploit historical natural experiments to more cleanly identify the causal effect of colleges on local invention than is possible in prior work. To investigate the channels through which this effect operates, I then link patent data from these natural experiments to a novel dataset of historical college yearbooks and to U.S. decennial population censuses. I find that most of the effect of colleges operates through indirect channels: the groups who are directly affected by a college, the college's alumni and faculty, account for only a small share of the patents in that college's county. Finally, I present a number of additional comparisons to provide suggestive evidence on the nature of these indirect channels.

Understanding which channels drive local invention is a topic of substantial practical importance. Academics and policymakers have reached a near-consensus that proximity to a college is a necessary condition to create a local invention hub. For example, Florida (2002b, p. 291-292) argues that "the presence of a major research university is a basic infrastructure component" of creative hubs, even more important than physical infrastructure like bridges and railroads. O'Mara (2005, p. 6) refers to colleges and universities as "the economic development engine" at the heart of innovative cities. And Moretti (2012, p. 186) concludes that "[a] research university was necessary but far from sufficient for the birth and coming-of-age of [Silicon] Valley." But if most of the local effect of colleges is driven by indirect channels not specific to the presence of a college, then other types of investments may be able to deliver similar local effects, calling into question the necessity of a college for invention hubs. I present evidence to this effect in this paper.

---

[1] Throughout this paper, I refer to all institutions of higher education as "colleges."

[2] See, for example, Moretti (2004) and Zucker, Darby, and Brewer (1998) for classic papers on the importance of college graduates and faculty in innovative communities, respectively.

I begin by estimating the local causal effect of establishing a college. Studying the causal effect of colleges is difficult because, as Hausman puts it: "To understand local industry effects of universities, one would ideally like to randomly allocate universities to locations and measure related industry activity in those locations after the universities arrived relative to before" (Hausman, 2017, p. 11). Conducting such an experiment is infeasible today. In this paper, I approximate this ideal experiment using historical data on the establishment of U.S. colleges from 1839 to 1954. By exploring the narrative record, I am able to identify "runner-up" sites that were strongly considered to become the site of a new college. The methodology is similar to that in Greenstone, Hornbeck, and Moretti (2010) to study the site selection decisions of large manufacturing plants.[3] The key idea behind this "runner-up" methodology is that, when selecting where to locate a new college, dozens of possible candidate locations are considered and iteratively eliminated; by the time only a few finalists sites are left they are likely similar along both observable and unobservable dimensions. While this methodology works well in the context of Greenstone et al. (2010), the identifying assumption can fail if only a small number of locations were ever considered, especially if these finalist locations are very dissimilar to one another. To account for this, I refine the methodology by restricting the sample to cases in which I can verify that the site selection decision is as good as random assignment.

A concrete example of a college site selection process is useful to fix ideas. In 1882, the state of North Dakota drew lots to determine where to locate the University of North Dakota and North Dakota State University; locations were literally randomly assigned (Geiger, 1958, p. 13-27). Not surprisingly, literal random assignment is rare, but many cases are approximately random. As an example of as-good-as-random assignment, in 1886 the citizens of Georgia wanted a technical college, but there was no consensus about where to put it. The two main rival sites were Atlanta and Macon. Both were known primarily as railway depots located in the interior of the state and were similar along a number of observable dimensions. A site selection committee assembled to vote on the location of the college. For the first 23 ballots, neither Atlanta nor Macon obtained the requisite majority. Finally, on the 24th ballot, Atlanta won over Macon by one vote (McMath Jr., Bayor, Brittain, Foster, Giebelhaus, and Reed, 1985, p. 24-32). It is thus easy to believe that Georgia Tech could have been located in Macon instead of Atlanta. The cases of the North Dakota universities and Georgia Tech were not isolated incidents: while the decisions were often less dramatic, these kinds of college site selection experiments occurred all across the United States during the second half of the nineteenth century and first half of the twentieth.[4]

---

[3] Helmers and Overman (2017) and Watzinger, Treber, and Schnitzer (2018) apply similar methodologies in other contexts.

The winning counties and the as-good-as-randomly-assigned runner-up counties are similar along observable dimensions. In contrast to my approach in this paper, most previous studies that focus on the establishment of new colleges assume that colleges are located at random (Andersson, Quigley, and Wilhelmsson, 2004,0; Cowan and Zinovyeva, 2013; Currie and Moretti, 2003; Frenette, 2009; Lehnert, Pfister, and Backes-Gellner, 2020; Moretti, 2004; Toivanen and Väänänen, 2016).[5] I show that assuming the random location of colleges overstates the effect of colleges on local invention.

I use patent data, assembled from a number of sources to fully cover the years 1836 to 2010, to proxy for local invention.[6] Using the runners-up as counterfactuals for the winning college counties in a differences-in-differences framework, I find that establishing a new college causes a sizable increase in local patenting. In the preferred specification, the college counties have about 48% more patents per year than the runner-up counties after the college is established. I argue that, if anything, this estimate likely understates the true local effect of establishing a college. I show that establishing a college also causes a roughly 49% increase in county population relative to the runner-up counties, providing the first suggestive evidence of the importance of indirect channels like population growth.

---

[4] See the historical appendix (Andrews, 2019c) for many more examples.

[5] Moretti (2004, p. 190-191), focusing exclusively on land grant colleges, writes that, "Land-grant colleges were often established in rural areas, and their location was not dependent on natural resources or other factors that could make an area wealthier. In fact, judged from today's point of view, the geographical location of land-grant colleges seems close to random."

[6] Of course, patent data are not a perfect proxy: not all inventions are patented, and not all patents are for meaningful inventions; see Griliches (1990) and Nagaoka, Motohashi, and Goto (2010) for a discussion of these issues. But patents have a number of attractive features as well. The patent data are available for the entire U.S. over a long time period and a wide range of technology fields, allowing me to estimate the long-run effects of establishing a diverse set of institutions. An additional benefit of patent data is that each patent document records the name of its inventor. By linking this to other datasets such as the population censuses, it is possible to discover the identity of the inventive individuals living in an area. I discuss this benefit in more detail below.

To explore the relative importance of different channels more directly, I match the names of patentees in college counties to other datasets to reveal inventors' identities. In particular, I match by inventor name to a novel dataset of historical college yearbooks and to U.S. population census data for the years 1940 and earlier to see if a patentee is an alumnus or faculty member of a particular college. By linking individuals across censuses, I can observe whether an individual was present in a college county at the time the college was established or appeared in the college county's census records after the fact. Through 1940, alumni and faculty of a particular college account for only about 5% of the patents in that college's county. More than 75% of patents are by individuals who were unaffiliated with their local college and did not live in the college county at the time the college was established. While the data on alumni and faculty can be linked to U.S. census data only through 1940, the decades I observe contain dramatic changes in technological development, urbanization, and migration. I therefore shed light on the role of colleges in driving these trends at the local level during an important period of both American economic history and the history of U.S. higher education. To get a rough sense of the extent to which the role of alumni and faculty have changed in recent years, I use data from Bell, Chetty, Jaravel, Petkova, and Reenen (2019) and Zolas, Goldschlag, Jarmin, Stephan, Owen-Smith, Rosen, Allen, Weinberg, and Lane (2015) to conduct a similar analysis for recent decades. I find that since 1996, in spite massive changes in the scale and scope of U.S. colleges, alumni and faculty of a particular college still account for less than 20% of patents in that college's county. It thus appears that indirect channels account for the vast majority of patents in a college's county, regardless of the time period investigated.

In the remainder of the paper, I conduct several additional analyses to provide suggestive evidence about the types of indirect channels that are most important. I test for the importance of positively selected migration using individuals' recorded occupations in the decennial censuses to determine if the migrants to college counties appear more skilled than the migrants to the runner-up counties. I find no difference in occupational income. Hence, positive selection does not appear to be the primary channel through which colleges indirectly affect local invention.

It may be the case that, even though the population in college and runner-up counties appear similar, non-college affiliated individuals near colleges benefit more from local knowledge spillovers than do individuals in runner-up counties because they can interact with educated faculty and alumni. To test this, I examine what I call "consolation prize" cases: instances in which the runner-up counties receive another type of state institution, such as a prison or insane asylum, instead of a college. The increase in patenting is statistically indistinguishable between the college and consolation prize counties. I also compare across different types of colleges.[7] Colleges that are focused on more technical fields might be expected to both produce larger direct effects and large spillovers to non-college affiliated individuals. I find no statistically significant difference across types of colleges, and if anything more technically focused colleges see smaller increases in local patenting. I view these results as further suggestive evidence that increasing population is the most important driver of local invention; spillovers resulting from this larger population do not depend on the activities going on at the college.

This paper is organized as follows. Section I describes the data, including an in depth explanation of the college site selection experiments. Section II presents the baseline results for the effect of the establishment of a new college on local patenting. In Section III I use the yearbooks and census data to show that only a small share of patents come from individuals directly affected by the college. Section IV further investigates the indirect channels through which colleges affect local invention. Section V concludes.

---

[7] Many studies of the effects of college focus only on one type of institution. For instance, Currie and Moretti (2003), Moretti (2004), and Kantor and Whalley (2019) all focus exclusively on land grant colleges. The ability to investigate various sorts of colleges is an additional strength of the analysis in this paper.

I Data and Empirical Model

I.A The College Site Selection Experiments

U.S. history provides an ideal laboratory to study the establishment of new colleges. As Goldin and Katz (1999) and Xiong and Zhao (2019) point out, the mid-19th to mid-20th centuries saw an explosion in the number of colleges and universities in the U.S., providing many potential college site selection experiments. In addition, because most colleges were built more than a century ago, it is possible to trace the long-term effects of colleges on invention. Moreover, many of these colleges were built at the beginning of what Goldin and Katz (2008) call the "human capital century" and Gordon (2016) refers to as "the golden age" of America's technological leadership.

To estimate the causal effect of establishing a college, the first step is to identify valid counterfactuals to college sites. I modify a methodology used by Greenstone et al. (2010) to estimate the agglomeration externalities from building large manufacturing plants, who use "runner-up" plant locations as counterfactuals for winning sites. I likewise use the historical record to find runner-up locations for colleges.[8]

A great deal of thought went into each college site selection decision throughout U.S. history. Horace Bushnell, a theologian who played a central role in locating both the University of California and the University of Illinois, articulated the weight of these decisions: "The site of a university is to be chosen but once. Once planted, it can never be removed; and if any mistake is made, that mistake rests on the institution as a burden to the end of time" (quoted in Ferrier (1930, p. 162)). Many localities wanted to secure a new college, and any economic benefits that went along with it, for themselves. This ensured that site selection decisions often became quite contentious. Further complicating the site selection decision is the fact that new colleges often had particular infrastructure needs. In the case of land grant universities, for example, the Morrill Act of 1862 explicitly prohibited states from using their land grant fund to construct buildings. This often forced states to locate land grant colleges in towns with unused buildings large enough for a college or in localities willing to raise the funds for construction. Thus, the set of potential candidate locations may be quite different from the average site within the same state.

---

[8] Patrick (2016) raises several challenges to the identification strategy employed by Greenstone et al. (2010). These critiques are unlikely to be valid in my context for a number of reasons. First, in contrast to for-profit firms, there is little strategic reason for colleges to hide their list of finalist locations from competitors. Second, I provide a great deal of institutional detail that shows that the site selection decision was indeed close to random. Finally, in Section I.F I show that college and runner-up sites are similar in terms of observables; in Appendix A.B I show that the colleges and runners-up evolved similarly as well.

To find runner-up sites, I consult detailed institutional histories, what Washburn (1979) calls "the driest of dry forms of historiography", for information on the college site selection process. I do this for 451 colleges, comprising almost 15% of the 3,039 degree-granting post-secondary schools currently in the U.S.[9] While it is not feasible to find and consult detailed histories for every college in the U.S. given the large number of small colleges, I attempt to find data on every prominent or influential U.S. school. More specifically, I investigate every national university ranked by the 2018 U.S. News and World Report Best Colleges ranking (`https://www.usnews.com/best-colleges/rankings/national-universities`), as well as the 25 best liberal arts colleges in the corresponding ranking (`https://www.usnews.com/best-colleges/rankings/national-liberal-arts-colleges`); every land grant college; the first public university founded in each state; the flagship university of a each state's public university system if this is different from either the land grant or first public university; every state technical school and mining college; every federal military academy; and every university belonging to a Power Five athletic conference. When data was available, I also investigated historically black colleges and universities (HBCUs) and private colleges, with a focus on the private colleges that have been historically noteworthy or are currently considered prestigious. For a handful of states, I also investigated each normal school established in that state. Over time, normal schools typically evolved to become "directional" state universities (for example, the Michigan State Normal College became Eastern Michigan University).

Of these 451 colleges, in 204 cases I am able to find information on the candidate locations that were considered.[10] Andrews (2019c) describes each of these 204 cases in much more detail. In Appendix A.A I provide additional comparisons of the sample of 204 colleges to the universe of U.S. colleges. In general, the colleges in my sample are larger than the average U.S. college in terms of number of students, faculty, library volumes, and other measures, and are thus also likely to have a larger than average effect on local invention and the local economy more generally, although they are similar on average to the typical "Carnegie R1 or R2" research university.[11]

---

[9] The count of U.S. colleges as of 2013 comes from `https://nces.ed.gov/fastfacts/display.asp?id=84`.

[10] For the other 247 colleges, I do not have any information on other candidate locations for the campus, either because histories are insufficiently detailed about the site selection process or, more commonly, because only one candidate location (the winner) was ever considered. The former occurs most often for private colleges when only a small number of decision makers were involved in picking a site, and hence the site selection process is often not transparent, as in the case of Johns Hopkins University. The latter occurs

One drawback to this approach is that it identifies all finalists, regardless of how similar the winning and losing sites are or how close the site selection process was to random assignment. For instance, three different counties submitted bids to the Ohio State legislature to receive the new Ohio State University, but there does not appear to be any serious discussion in the legislature: the college was always intended to be located at the state capital in Columbus. Moreover, Columbus is very different from these other localities along observable dimensions. To mitigate this problem, I further restrict the sample to only include cases in which, conditional on being a finalist, the site selection decision is as good as random; I refer to these as "high quality" college selection experiments. I consider 75 of the college cases to be high quality experiments. Eight of these high quality experiments take place prior to the start of the patent data in 1836, so I exclude them. The remaining 67 high quality college site selection experiments form the baseline sample.

While each of the high quality site selection experiments is unique, they can broadly be grouped into four categories. First, a vote among candidate locations may be exceptionally close; the case of Georgia Tech described in the Introduction is one example of this. Second, candidate locations frequently submitted bids to boards of trustees or state legislatures to receive a new college. When two bids are similar, this is evidence that the localities valued receiving the school roughly equally, and the decision makers were largely indifferent between

---

frequently for universities located in large cities; cities often decided to establish their own colleges once they reached a threshold size, with no consideration of other places to locate the school. This occurred, for instance, in the cases of the University of Louisville, Wichita State University, and Loyola University Chicago.

[11] I emphasize that while the sample colleges are larger and more prestigious than average, it is not possible to identify runner-up sites for the extreme right tail of U.S. colleges, such as Harvard, MIT, and Cal Tech. For policymakers looking to use a college as a tool to promote local invention, the "average" U.S. college is arguably a more realistic treatment than one of the world's most elite institutions. In Appendix F.C, I discuss heterogeneity by the quality of a college.

the two sites. Third, in some instances a new college had specific infrastructure needs, such as existing vacant buildings of a suitable size, and only two or three such sites within the state possessed the required infrastructure. Finally, some site selection experiments involve quirky random events that are difficult to otherwise classify. The random assignment of the University of North Dakota and North Dakota State University is an example of this. Cornell University provides another example. Ezra Cornell and Andrew White, the fathers of Cornell, wanted to establish the college in one of their home towns but could not decide on which. Ezra Cornell was from Ithaca, while Andrew White was from Syracuse. But Cornell had been cheated of his wages as a young man in Syracuse and refused to locate the college there. Consequently, Cornell University is located in Ithaca.[12] In Appendix D, I show that the results are insensitive to discarding any of these groups of high quality experiments. The results are also not sensitive to reclassifying marginal cases as either high or low quality.

Augmenting the runner-up methodology by using narrative history to exclude cases in which selection is not as good as random assignment is, as far as I know, novel in the literature.[13] In a paper studying agricultural experiment stations, Kantor and Whalley (2019) compare a subset of land grant colleges to all contending locations as a robustness check. They do not, however, restrict attention to those cases in which the winning site is as good as randomly assigned; for instance, Ohio State is one of the colleges in their sample. In the analysis below, I show that failing to exclude these low quality experiments overstates the effect of establishing a college.

---

[12] I have been unable to find any evidence that Syracuse tended to have citizens of a lower moral character than did Ithaca. Syracuse and Ithaca were furthermore similar along observable dimensions before the establishment of Cornell University. Syracuse would, of course, get its own university several years later. In Appendix D.B, I discuss several different strategies to handle runner-up sites like Syracuse. In general, to the extent runner-up counties are likely to receive a college of their own in the future, the results in this paper should be interpreted as a lower bound on the local effects of college.

[13] Liu (2015), Bonander, Jakobsson, Podestà, and Svensson (2016), and Lee (2019) use synthetic control methodologies to study the economic impact of establishing or expanding colleges. A synthetic control methodology is less appropriate in a historical context because data for several desired predictors are not available for most locations in most pre-treatment years. For instance, Lee (2019) argues that real estate prices are an important predictor to understand the demand for land in winning and losing locations. In addition, in most cases unobservable factors, such as the enthusiasm of the local population for education or the presence of specific pieces of infrastructure, were crucial both in becoming a finalist site as well as in the production of innovations. These factors are taken into account in the current methodology but are neglected in any methodology that matches on observables.

I.B Patent Data

Patent data extend from the first year for which U.S. patent records are complete, 1836, to 2010. The patent data come from four sources, with different sources available for different years. For the years 1836-1870, I use patent data collected in the Subject-Matter Index of Patents for Inventions Issued by the United States Patent Office from 1790 to 1873 (Leggett, 1874), compiled by Dr. Jim Shaw of Hutchinson, KS. I use the Annual Reports of the Commissioner of Patents for the years 1870 to 1942. See Sarada, Andrews, and Ziebarth (2019) for details on cleaning, parsing, and preparing this dataset. The years 1942 to 1975 come from the HistPat dataset compiled by Petralia, Balland, and Rigby (2016a); see Petralia, Balland, and Rigby (2016b) for details on the construction of this data. Finally, for the years 1975 to 2010, I use the patent data created by Li, Lai, D'Amour, Doolin, Sun, Torvik, Yu, and Fleming (2014) which contains cleaned inventor names. Because all analyses include year effects, there is no concern with the fact that different years make use of different patent data sources.[14] Each of these datasets contains, for every granted U.S. patent, the names and residence of all inventors. I aggregate patenting to the county level.

I merge by patent number and/or inventor name to other datasets that include additional patent information. The U.S. Patent and Trademark Office's Historical Patent Data Files (Marco, Carley, Jackson, and Myers, 2015) contain information on patent classes for historical patents. Enrico Berkes graciously provided data on patent citations and patent claims; see Berkes (2018) for details. Andrews (2019a) contains a detailed discussion of all of these patent datasets.

I.C Yearbook Data

College yearbooks are available from `ancestry.com`. The college yearbooks include full student names, which can be used to match students from yearbooks to other data sources such as the patent record or the US decennial censuses. The yearbooks usually include names of faculty members as well. I collect yearbooks from 20 different colleges, roughly 30% of the colleges in my sample, covering 305 yearbooks from 1879 to 1940 and including records for 83,448 undergraduate seniors and 30,541 faculty members.[15] The yearbook data, including a discussion of how representative the yearbook colleges are relative to the rest of the sample, are described in more detail in Appendix C.A.

---

[14] The use of different patent datasets in different years could still bias the results if, for instance, one dataset systematically recorded more patents from college counties than runner-up counties. Because the college and runner-up counties are so similar to one another as described in Section I.F below, this concern is minimal.

I.D Matching Patent and Yearbook Data to the U.S. Census Data

I use the names of individuals in the patent and yearbook data, along with the 100% U.S. federal decennial population censuses, to determine which patents in a college's county were granted to alumni and faculty of that college. The census data for the years 1850-1940 are transcribed by `ancestry.com` and the Minnesota Population Center and hosted by the NBER.[16]

The yearbook data provide the flow of students passing through each college, while I am interested in the stock of each college's alumni. For each year, I therefore compile names of college seniors from the previous yearbook years to assemble a stock of known alumni names. Once I have data on the stock of alumni for each year, I fuzzy match the names of alumni, faculty, and patent records for each county i and year t to the census records for county i in the nearest census year.[17] For the non-alumni, non-faculty patents, I am interested in determining whether these individuals were present in county i at the time college i was established, or whether they first appear in county i later. To do this, I link records from county i in a given census year to county i's records in the previous census; by doing this for all censuses, I can document the earliest census for which each individual appears in county i.[18] If an individual is neither an alumnus nor faculty and first appears in county i before college i is established, I refer to the individual as a "pre-college other," while if he first appears after college i is established I refer to him as a "post-college other."

---

[15] These need not be *unique* faculty members; in most cases the same faculty member is listed in multiple yearbook years. In rare cases, yearbooks do not record names of college seniors but do list names of juniors. In these cases, I record the names of juniors instead.

[16] No individual-level census data exists for the 1890 census, which was destroyed in a fire. Aggregate county-level data from the censuses are available for 1840 and 1890, as well as for later census years, but individual-level data are not available.

[17] So, for example, I match patent records from Grand Forks County, ND, and yearbook records for the University of North Dakota (located in Grand Forks) from 1935-1940 to the 1940 census records for Grand Forks County, and 1930-1934 patent and yearbook records to the 1930 census records for Grand Forks County. Throughout, I only use census records for males because name changes by females produce artificially low match rates.

[18] More specifically, I link records across censuses following the procedure in Ferrie (1996). I consider an individual j in census year t to match to an individual k in the prior census year $t-10$ if the two individuals

An adjustment to the stock of college alumni is necessitated by the fact that, for a given college, yearbooks may not be available for all consecutive years. When this occurs for college i, the calculated stock of alumni will provide an undercount of the true number of college i's alumni; when yearbooks are missing for many years, this undercount can be substantial. To overcome this, for each college i I calculate a patenting rate for college i's alumni by dividing the number of alumni matched to patents in the census for county i and year t by the number of alumni matched to the census for county i and year t, averaged over all years. Next, I interpolate the number of college seniors for the missing yearbook years and use these interpolated counts to reconstruct the stock of alumni as above. I apply the patenting rate for college i's matched alumni to the adjusted stock of college i's alumni to get the adjusted number of patents belonging to alumni of college i. I correspondingly scale down the number of pre-college others and post-college others in county i by the number of additional alumni patents resulting from this interpolation procedure, keeping the relative sizes of the pre- and post-college others the same.

This procedure gives me the number of patents in county i belonging to the alumni and faculty of college i, along with the number of patents by individuals who were first in county i before and after college i was established. Appendix C.B describes this procedure in more detail. Appendix E.A discusses the sensitivity of the final alumni and faculty patenting shares to alternative interpolation assumptions and matching criteria. The alumni, faculty, pre-college others, and post-college others categories are constructed to be exhaustive and mutually exclusive. As Bailey, Cole, Henderson, and Massey (2018) show, a simple fuzzy matching procedure of individuals' names across datasets, as I use here, can produce a large number of false positive matches. Thus, because I count as a match any name that is similar across the patents, yearbooks, and census records, these results likely overstate the share of patents belonging to alumni and faculty. I use this sample of matched patent-yearbook-

---

have a sufficiently similar name, if the race and gender are the same ($Race_j = Race_k$, $Gender_j = Gender_k$), if the birthplace is the same ($Birthplace_j = Birthplace_k$), if mother and father birthplace are the same ($MotherBirthplace_j = MotherBirthplace_k$, $FatherBirthplace_j = FatherBirthplace_k$), and if $age_{kt} + 2 \geq age_{jt} - 10 \geq age_{kt} - 2$.

census data for the results in Section III.

I.E County Data

County-level data comes from the National Historic Geographic Information System (NHGIS) (Manson, Schroeder, Riper, and Ruggles, 2017). The NHGIS data allows me to compare counties along a number of useful dimensions including population; composition of the county population along racial, gender, immigration, and age dimensions; urbanization; and average wages and production in both agricultural and manufacturing sectors. I also use data on the total number of accredited colleges at the county level. These are found in Reports of the Commissioner of Education, several years of which have been transcribed: 1870, 1875, 1880, 1885, 1890, 1895, 1900, 1905, 1910, and 1914 by Heyu Xiong and Yiling Zhao; and 1897, 1924, and 1934 by Claudia Goldin.

I.F The College and Runner-Up Counties

Appendix Table A1 lists each of the 64 high quality college site selection experiments in the final sample as well as the year in which the experiment took place. Table 1 further summarizes the college experiment data. Each college site had on average 2 runner-up sites. The runner-up sites are on average about 140 km ($\approx$87 miles) away from the college towns, with the median runner-up about 90 km ($\approx$56 miles) away, using geodetic distances. This is far enough that the college and runner-up sites are typically in different labor markets, but close enough to be affected similarly by region-wide shocks. Figure 1 is a map of the college and runner-up counties throughout the U.S., providing visual verification that the college and runner-up counties vary in their distance from one another. The map also shows that the entire continental U.S. is represented in the sample and that colleges were not simply built near existing major population centers. Row 3 of Table 1 shows that colleges were established throughout the entire period from 1839 to 1954, with the mean and median college established in the mid-1880s. The median college began admitting students four years after determining where the school would be located. To give a sense of the type of colleges involved in the study, I classify colleges into one of seven mutually exclusive groups: land grant colleges, technical colleges, normal schools, historically black colleges and universities (HBCUs), military academies, other public colleges, and other private colleges.[19] A plurality of the college experiments involve land grant colleges. 10% of the experiments involve technical colleges, 17% involve normal schools, 4% involve HBCUs, and 4% involve military academies. 17% of the colleges are classified as "other" public colleges, while 7% are classified as "other" private colleges.

Figure 2 compares the college and runner-up counties and shows that the runners-up are a better match for the college counties than are the "non-experimental" counties, which are all other counties in a college's state that are neither college nor runner-up counties. The black diamonds display the difference in the mean between the college and runner-up counties in the last U.S. census year before the college was established on a number of economic, demographic, and educational variables, expressed as a fraction of the mean value in the college county. The black lines show 95% confidence intervals of a simple t-test of the difference in means. I use census years because most of the demographic and economic variables are collected with the decennial census. The means of the college and runner-up counties are statistically indistinguishable and remarkably similar in magnitude. For readability, I winsorize the top 1% for all variables; the magnitude and statistical significance of the differences in means are nearly identical when using non-winsorized data. The green circles show the difference in the mean between the college and the non-experimental counties, which are the counties in each state that are not classified as either college or runner-up counties. The green dashed lines show 95% confidence intervals for the t-test. The college and non-experimental counties also tend to be similar along several dimensions, making Moretti's (2004) claim that colleges were located "close to random" understandable. But relative to the non-experimental counties, the college counties tend to have a statistically larger population and a larger share of the population living in urban areas; the differences in means between the non-experimental and college counties is typically larger than the differences between runner-up and college counties for other characteristics as well, although not statistically different from zero. Appendix A.B provides even more evidence that the college and runner-up counties are similar to one another prior to the establishment of a new college.

---

[19] Technical colleges include schools focused on engineering, mining, and industrial arts. Normal schools are colleges focused on teacher training; many of these have evolved to become directional state universities. Other public and private universities include all public and private, respectively, schools that do not fit into any of the other classifications. For instance, the University of Texas is classified as an "other public" college in the sample; Texas also has two other state-wide (that is, not "directional states" targeted to a particular region within Texas) public universities, a land grant college (Texas A&M) and a technical college (Texas Tech), both of which are also in my sample. In some cases, a college may fall into multiple categories. For example, many HBCUs are also state land grant colleges. For clarity, in Appendix Table A1, I place each college into its "best" category. All results are insensitive to reclassifying colleges.

I.G Empirical Model

I estimate straightforward differences-in-differences model. In county i at time t, the number of patents is given by

$$PatentMeasure_{it} = \delta_1 College_i * PostCollege_{it} + \delta_2 PostCollege_{it} + County_i + Year_t + \epsilon_{it}, \tag{1}$$

where $College_i$ is an indicator variable equal to one if county i receives a college, $PostCollege_{it}$ is an indicator variable equal to one in years t after i's college has been established, $County_i$ is a county fixed effect, $Year_t$ is a year effect, and ϵ_{it} is a county-year varying error term. With only a single college site selection experiment, the term $PostCollege_{it}$ would be redundant because the post-college dummy is perfectly co-linear with the year effects. There are multiple experiments in the dataset, however, with each college being established in different years, and so each set of counties will be in the post-college period in different years. The year effects therefore control for nationwide time-variant changes in patenting, while $PostCollege_{it}$ controls for changes that occur within all counties affiliated with a given college experiment after that experiment occurs.

II Baseline Results

Figure 3 plots smoothed residualized logged patenting for college, runner-up, and non-experimental counties separately.[20] The year in which a new college is established is normalized to be year 0 for all experiments. Three results are immediately clear. First, new colleges do not appear to be randomly located; there is a large difference between the college and runner-up counties on one hand and the non-experimental counties on the other, both in the level and growth rate of patenting. Second, the college and runner-up counties patented similarly in pre-college years, providing suggestive evidence that the experimental design is valid. Third, after the establishment of a new college, the college and runner-up counties diverge, with college counties patenting more. This divergence is especially pronounced after about half a century.

---

[20] Figure 3 is constructed by regressing $\log(NumPat_{it} + 1)$ on year effects $Year_t$ and then plotting the residuals using local mean smoothing with an Epanechnikov kernel function. Removing time effects is useful

Table 2 formalizes the intuition in Figure 3.[21] The columns show different regression specifications. The coefficient of interest is displayed in the first row and shows the estimated proportional change in the number of patents generated in the college county relative to the runner-up county after establishing the new college. For all columns, standard errors are clustered at the county level.[22]

Column 1 shows the results of estimating Equation (1) where the dependent variable is $\log(Num.Pat_{it}+1)$. College counties have about 48% more patents per year than runner-up counties. The average county had about 4.8 patents in 1880, around the time the median new college is established. This translates to just over two additional patents in 1880. By 2010, the average county had about 37.5 patents per year, so the college causes almost 17 additional patents per year in 2010. This result is statistically significant at the 1% level.

In Column 2, I use the inverse hyperbolic sin of patents as the dependent variable.[23] The benefit of the inverse hyperbolic sin is that it can take on zero values, yet still has the same simple interpretation as in Column 1. Here, establishing a new college causes 57% more

---

because, as Griliches (1990) shows, there has been a secular increase in patenting over time as well as country-wide cyclical fluctuations in patenting that coincide with business cycles and changes in the administration of the Patent Office; failure to control for these factors makes interpreting the graph more difficult. The figure contains a balanced set of college experiments, including only counties with at least 20 years of pre-college and 80 years of post-college data available. The graph is similar when using an unbalanced panel instead.

[21] Appendix D.A explores the dynamics observed in Figure 3 in much more detail.

[22] I also cluster at the state, experiment, state×year, and experiment×year levels. I additionally cluster at multiple levels as proposed in Cameron, Gelbach, and Miller (2011): I cluster at the county and year; state and year; experiment and year; and county, state, experiment, and year levels. Clustering at the county level produces the largest standard errors, but the standard errors are virtually identical at every level and none of the inferences change.

[23] The inverse hyperbolic sin of patents is given by $\text{arcsinh}(NumPat_{it}) = \log(NumPat_{it} + (NumPat_{it}^2 + 1)^{\frac{1}{2}})$.

patents per year in the college counties relative to the runners-up, similar in magnitude to the results in Column 1.

Column 3 shows results using an alternative calculation of logged patents as proposed by Blundell, Griffith, and Van Reenen (1995). Rather than adding a positive constant before taking the log of patents, this alternative method uses $\log(NumPat_{it})$ as the dependent variable. Whenever $NumPat_{it} = 0$, a dummy variable is set to one and $\log(0)$ is replaced with 0. In this specification, establishing a new college leads to a roughly 27% more patents per year.

Column 4 uses the fact that the number of patents takes on integer values and presents estimates of a Poisson regression, expressed as incident rate ratios. Because of the strong influence of outliers when raw counts of patents are used, I Winsorize the top 5% of counties by yearly patenting. In this specification, establishing a new college leads to an 89% increase in patenting, about twice as large as the baseline estimate in Column 1.

In all specifications, establishing a new college causes a sizable increase in local patenting. How to interpret these estimates? First, I stress that the design used in this paper does not estimate the effect of establishing a college at a random location. Rather, I estimate the local effect of establishing a college in a county conditional on that county being considered a good location for a college; in essence this is a local average treatment effect for the subset of U.S. counties suitable for a college. For that subset of counties, the sizable estimated magnitude may actually be an underestimate if runner-up counties are themselves likely to get a college of their own at a later date. I explore this possibility in more detail in Appendix D.B and find that 60-75% of runner-up counties receive a college of their own at some point, depending on how one defines a college. Excluding these cases can result in estimates about 50% larger than the baseline estimates, and so the results in Table 2 should be interpreted as a lower bound of the local effects of establishing a college for the set of suitable counties.

In Section III below, I document the importance of indirect effects by matching patents to yearbook records for a subsample of college experiments for which yearbooks are available. To better understand this yearbook this sample of "yearbook colleges," in Column 5 I repeat the baseline specification from Column 1 but use only the sample of yearbook colleges. Establishing a yearbook college causes 81% more patents per year in the yearbook college counties relative to their runner-up counties, so the yearbook colleges appear to have a somewhat larger effect than the full sample of colleges. I present several additional specifications using only the yearbook data in Appendix C.D.

II.A Robustness Checks and Extensions

Appendix D presents numerous robustness checks. In particular, I show that the baseline results are robust to a battery of additional specifications (Appendix D.C), to alternative ways of handling counties with other local colleges at the time of treatment (Appendix D.D), to aggregating the data at different geographic levels (Appendix D.E), and to using different subsets of the college site selection experiments (Appendix D.F). I also present results from placebo tests that use other years as the "experiment date" and find no effect on patenting (Appendix D.G). I further verify that the results are not driven by a shift in the types of inventions patented (Appendix D.H) or by patent quality (Appendix D.I) after establishing a new college.

In Appendix D.J, I repeat the baseline results in Table 2 but instead of using only the high quality college site selection experiments, I use all runner-up sites, including the cases in which the winning site was not as-good-as-randomly assigned. While qualitatively similar, the magnitude of the estimated effects are larger, suggesting that there was positive selection of the college sites when the location was not decided essentially at random.

Finally, Andrews (2019b) uses the same college site selection experiments to show that the results are not driven by migration from the runner-up counties to the college counties.

II.B Colleges and Population

One simple channel through which colleges may mechanically increase local patenting is by increasing the local population. Indeed, several studies highlight the role of colleges in attracting individuals and firms to an area and inducing urbanization (Cantoni and Yuchtman, 2014; Drucker and Goldstein, 2007; Florida, 2002b; Holley and Harris, 2016).

In Column 6 of Table 2, I show that establishing a new college does indeed cause an increase in county population relative to the runner-up counties. Because county population is only collected from the decennial U.S. population censuses, the sample in Column 6 is restricted to census years. Population exhibits a nearly identical percentage increase to patenting: establishing a new college increases population by about 49% in the college counties relative to the runners-up, and patenting by 48%. This finding is the first, albeit highly speculative, evidence that indirect channels in general and population increases in particular are one of the most important ways through which colleges affect local patenting.[24] In the following sections, I present further evidence that indirect effects account for the majority of patenting in college counties. Appendix D.K presents additional results on colleges and population.

III Who Do the Patents Come From?

To more directly analyze the channels through which colleges affect local invention, I set aside the runner-up sites for now and instead examine the identities of patentees within college counties. The prior literature has struggled to convincingly establish the channels through which colleges promote local patenting. Dating to Jaffe (1989), the literature typically observes how patenting by firms co-located with a university co-vary with university activity; if any change in firm patenting is observed, this is counted as a "knowledge spillover" from the university to these firms, implicitly attributing to indirect channels inventions that may be the result of direct channels. The issue, as noted by Zucker, Darby, and Armstrong (1998), Almeida and Kogut (1999), Breschi and Lissoni (2001), Breschi and Lissoni (2009), and Leten, Landoni, and Looy (2014), is that nearby patents may be coming from individuals who are directly affiliated with the college as either alumni or faculty. Data issues usually prevent researchers from discovering which individuals have a direct affiliation with a college and which do not.[25]

To overcome this problem, I use the linked yearbook-patent-census data described in

---

[24] Of course, there are other ways to interpret this result. For instance, increases in patenting may induce more migration to a county. Alternatively, the fact that the magnitudes of the estimated effect on patenting and population are similar may be purely coincidental.

[25] Several recent papers document that college-educated individuals are more likely to invent (Aghion, Akcigit, Hyytinen, and Toivanen, 2017; Akcigit, Grigsby, and Nicholas, 2017b; Bell et al., 2019; Jung and Ejermo, 2014). Maloney and Caicedo (2017) argue that the skilled engineers trained at U.S. colleges are a key source of local innovation. None of these studies can link a college graduate to a *particular* college.

Section I.D to show the share of patents in college counties coming from individuals who are directly affiliated with a college and the share who are only indirectly affected. As these results only rely on the composition of inventors within college counties, they do not depend on the choice of runner-up counties used in the baseline analysis in Section II.

Table 3 shows population and patenting for a college's alumni, a college's faculty, and non-college-affiliated individuals who were first present in the college county before and after a college is established ("pre-college others" and "post-college others," respectively) for the 20 college counties for which yearbook data is available. In Columns 1 and 2, I list the number of linked male individuals and the share of the male county population belonging to each group in the yearbook college counties. In Columns 3 and 4, I list the number of patents and share of census-matched county patents belonging to each group. Finally, in Column 5, I show the average patenting rate of each group, calculated as $Num.Patents_g * 10,000/Num.Members_g$ for members of group g that match to the census in the college county.

On average, a college's alumni account for only about 5.3% of all patents in the college county. Faculty account for a negligible share of all patents, so together the two groups directly affiliated with the college account for about 5% of all patents in the yearbook college county. As mentioned in Section I.D, these results likely overstate patenting by alumni and faculty; I show in Appendix E.A that alternative approaches to attributing patents to alumni and faculty result in an even smaller share of patents belonging to those groups.

I stress that these results do not show that alumni are unlikely to patent, but rather that alumni do not account for many patents in the counties where they obtained their degrees. Even if alumni patented at very high rates (Column 5 suggests this is unlikely), there are simply too few to meaningfully change overall patenting in the college's county, likely due to out-migration after graduation.[26] Several studies emphasize that college graduates tend to be highly mobile (Bound, Groen, Gézdi, and Turner, 2004; Groen, 2004; Sumell, Stephan, and Adams, 2008; Zolas et al., 2015), so it may be difficult for a college's county to retain its highly inventive alumni.

---

[26] I also caution that these results do not suggest that college attendance caused the alumni inventors to patent; the alumni patentees may have been likely to invent even in the absence of a college. The causal relationship between college attendance and patenting is ambiguous; for instance, Bianchi and Giorcelli (2019) argue that in some cases attending college may even cause talented individuals to go into careers like public administration that patent at low rates. Understanding the causal effect of college attendance on individuals' propensity to patent is an important topic for future work.

III.A Is the Role of Alumni and Faculty Similar Today?

It is possible that patenting and migration behavior for alumni and faculty are very different today than they were in the decades immediately after colleges were established. In particular, several studies show that the number of patents assigned to colleges began increasing in the post-World War II era, and then increased dramatically following the passage of the Bayh-Dole Act in 1980 (Arora, Belenzon, and Lee, 2019; Mowery and Sampat, 2001; Mowery and Ziedonis, 2002; Sampat, 2006). Likewise, Figure 3 suggests that colleges have a much larger effect on local invention after the college has been established for several decades. The transcribed yearbook data used in Table 3 only cover years through 1940, making it difficult to assess the external validity of these results for current innovation policy.

In this section, I present suggestive evidence that alumni and faculty account for a modest share of patents in college counties in recent years, and hence the conclusions from Table 3 likely remain largely valid today. To do this, I create proxies for alumni and faculty patenting. I describe the construction of these proxies and the results in more detail in Appendix E.C.

Bell et al. (2019) match college graduate inventors to their alma maters and compute the number of patents granted from 1996-2014 to alumni of cohorts born from 1980-1984. Because I am interested in the number of patents belonging to all alumni of a particular college (rather than just the patents from a particular birth cohort), I use data from the Integrated Postsecondary Education Data System (IPEDS) to construct alumni counts for each of the yearbook colleges.[27] I then multiply this stock of alumni by the patenting rate for the 1980-1984 birth cohort.[28] This gives the number of patents produced by the alumni of the yearbook colleges while residing anywhere in the U.S.; the results in Table 3, on the other hand, document the share of patents belonging to the alumni who are residing in the college county. To determine the role of alumni on local invention in recent years, I scale the number of alumni patents by the probability that college alumni reside close to their alma maters calculated in Zolas et al. (2015). After making these adjustments, I find that alumni can explain 14.5% of the patents in the college county over the period 1996-2014.

---

[27] See `https://nces.ed.gov/ipeds/` for the IPEDS data.

[28] If the alumni patenting rate is increasing over cohorts, then applying the alumni patenting rate for the 1980-1984 birth cohort, which would have attended college between 1998 and 2006, to earlier birth cohorts will overstate the number of patents attributed to alumni. The lifetime patenting behavior of the 1980-1984 birth cohort is also necessarily truncated; members of this cohort are at most 35 years old during the study period, which is below the average age of all inventors calculated by Bell et al. (2019) for the 1996-2014 time period. It is unclear how this will bias the empirical patenting rate.

I likewise do not have a direct measure of faculty patenting in recent decades, so I construct counts of university-assigned patents. This is the same measure used in the work on the Bayh-Dole Act. The number of university-cited patents grows dramatically in recent decades for the sample of yearbook colleges, consistent with patterns others document. But overall patenting in these college counties also grows rapidly, so that the share of university-assigned patents is only about 4.5% of all patents over the period 1996-2014.

Together, alumni and faculty therefore account for about 19% of all patents in a college's county.[29] From these back-of-the-envelop calculations, patenting by alumni and faculty appears to have increased over time; today the share of alumni and faculty patents is about 3.5 times larger than in the pre-1940 years. But even for recent decades, alumni and faculty still likely account for less than one in five patents in their college counties.

IV Further Investigating the Indirect Channels

IV.A Are Migrants More Positively Selected in College Counties?

By far the largest share of patents in Table 3 comes from individuals who were not living in the college county at the time the college was established. This group accounts for more than 75% of county patenting on average, and just under 65% of county population. One possibility, as argued in, for instance, Florida (2002a) and Florida (2002b), is that colleges are extremely attractive local amenities that may attract particularly skilled and inventive individuals. This sounds especially plausible in light of findings that inventors are particularly geographically mobile and often migrate to be close to other inventors (Aghion et al., 2009; Akcigit, Baslandze, and Stantcheva, 2016; Akcigit, Grigsby, and Nicholas, 2017a;

---

[29] Note that this double counts some alumni and faculty patenting if alumni assign their patents to the college for any reason, for instance if the patent is based on work completed in a university lab while they were a student.

Kerr and Lincoln, 2010; Moretti and Wilson, 2014,1; Waldinger, 2016). These "post-college others" patent at a higher rate than any other group, lending prima facie evidence that positively selected mobility may be at work.

To test whether colleges are driving the migration of particularly skilled individuals, it is necessary to see whether these post-college others are more positively selected in the college counties than in the runner-up counties. While information about each individuals' wages and education level are not recorded or transcribed in most historical censuses, occupations are recorded for the census years 1900-1940. I follow the literature and use the average income of each occupation in 1950 to create an average occupational income score.[30] Since most colleges were already established by the time the individual-level occupation data comes available, a differences-in-differences approach is not applicable. Given that the college and runner-up counties are similar to one another in the pre-college years for all observable variables, there is no reason to expect occupational income to differ in a systematic way between them prior to establishing a new college. In Table 4 I therefore show first differences, comparing average logged occupational income in the college to runner-up counties, including year and experiment fixed effects and clustering standard errors at the county level.

Column 1 compares average income for all individuals in the college counties compared to the runner-up counties. Column 2 compares all individuals who were not present in the college or runner-up counties at the time the college was established, linking all individuals to prior censuses as described in Section I.D for the yearbook counties. Columns 3 and 4 repeat Columns 1 and 2 but restrict attention to the patentees. In all cases, the average income in college counties is statistically indistinguishable from that in runner-up counties, and if anything individuals in the runner-up counties perhaps have higher occupational incomes.[31]

Looking at average occupational incomes may be missing changes in the composition of occupations in college counties relative to runner-up counties. In particular, college counties may be relatively more successful at creating an environment suitable for innovative sectors. To check this, I compile a list of the most common inventor occupations for the decades

---

[30] See, for instance, Feigenbaum (2018) for a discussion of the strengths and weaknesses of the commonly-used occupational income scores.

[31] Results in all columns are calculated conditional on having non-missing occupation data for individuals in these groups. Because some counties have no patents in a given year, and in some counties occupational data is missing for all individuals of a given type, the number of observations varies across regressions.

1900-1940 based on all inventors matched to the U.S. censuses in Sarada et al. (2019); Appendix F.A provides more details on these common inventor occupations. In Column 5, I show that the share of individuals belonging to a common inventor occupation is similar in the college and runner-up counties.

In sum, while establishing a college increases population and hence patenting, they do not appear to be attracting individuals who have observably greater skills than the individuals in the runner-up counties or creating relatively larger innovative sectors, at least for the counties and years for which occupation data is available.

IV.B "Consolation Prizes"

Many indirect channels, such as attracting population or inducing agglomeration economies, may not be unique to colleges. On the other hand, if living close to faculty members or students induces large knowledge spillovers, then colleges may be playing an essential role for local invention even if alumni and faculty do not account for many patents. In this section, I test the null hypothesis that these college-specific indirect channels are negligible. In particular, I compare counties that receive colleges to counties that receive what I refer to as "consolation prize" institutions at the same time. In these cases, states typically allocated several institutions at the same time, including the state capital, the state prison, or the state insane asylum. While numerous localities may have been lobbying to get a state institution, which locality ended up with which institution was as good as random. These consolation prize cases are especially common in western states that were largely unsettled and achieved statehood after the passage of the Morrill Act in 1862. In one famous example, the Tucson delegation set out for Prescott for the Arizona territorial legislature in 1885 intent on getting the state mental hospital. But flooding on the Salt River delayed the delegation. By the time they reached Prescott, the mental hospital had already been spoken for; Tucson was stuck with the state university.[32]

Because a consolation prize experiment is only possible when a state (rather than a private organization or regional group) is deciding the location of several institutions at once, in Column 1 of Table 5, I repeat the baseline estimates from Table 2 but include only the subset of experiments that are "potential consolation prize" experiments, namely any experiments in which a land grant, technical, or other public college was established and designed to serve the needs of the entire state; in other words, I exclude private colleges and regional public colleges. The estimated coefficient is similar to, although a bit larger than, the baseline estimate with the entire sample. Column 2 shows results that compare college counties to only the consolation prize runner-up counties. I am unable to reject the null hypothesis that patenting in college and consolation prize counties is identical after establishing the new college and, moreover, the coefficient is close to zero in magnitude: college counties do not appear to cause more patenting than counties that receive prisons, hospitals, or insane asylums. Panel (a) in Figure 4 presents these results graphically, analogously to Figure 3, and shows that the college and consolation prize counties evolve remarkably similarly over many decades. Panel (b) suggests why this may be the case: population in the consolation prize counties grows nearly identically to population in the college counties. Instead of repelling highly mobile workers, as prisons or asylums might today, the consolation prizes gave small towns an identity and attracted people and firms to the area. Column 3 repeats Column 1 but excludes the runner-up counties that receive a consolation prize.[33] Now, a new college increases patenting by about 61%, larger than the 50% baseline estimate. None of these results are qualitatively changed by excluding any particular type of consolation prize county, in particular those that might attract high human capital individuals such as state capitals, nor by excluding consolation prize counties that later establish a college of their own; see Appendix F.B.

---

[32] For the purposes of this exercise, it does not matter which institution was the "consolation" and which was the "prize." For more details on the site selection decision of the University of Arizona, see Martin (1960, p. 21-25), Wagoner (1970, p. 194-222), and Cline (1983, p. 2-4).

IV.C College Types

I next examine variation across types of colleges. If spillovers from university-affiliated individuals with advanced technically educations are especially large, or if the direct channels of alumni human capital and faculty research are the primary ways by which colleges promote invention, then colleges that focus on technical skills should cause much larger increases in local invention than other types of colleges. To test this, I classify colleges by type as described in Section I.F. From these college types, I further classify each college as either a "practical" or a "classical" college. Practical colleges are land grant colleges or technical schools. Classical colleges are normal schools and other private and public colleges. Land grant colleges were required by law to provide instruction on "agricultural and mechanical arts", and technical colleges explicitly focused on skills such as engineering, mining, or industry. At the same time, normal schools trained public school teachers, and so typically devoted less, if any, attention to technical skills. Other public and private colleges tended to have a less practical focus, providing instruction in classes like law, religion, or languages.[34]

---

[33] Because in some experiments all of the runner-up counties received a consolation prize, the number of experiments is smaller in Column 3 than in Column 1.

I interact a dummy for classical and practical colleges with the $College * PostCollege$ and $PostCollege$ terms. The results are presented in Column 4 of Table 5. For readability, I only display the coefficients for the $PracticalCollege * College * PostCollege$ and $ClassicalCollege * College * PostCollege$ interaction terms. If anything, the classical colleges cause a larger increase in local patenting than do the practical colleges (68% more patents per year in the classical college counties relative to their runners-up compared 31% more for the practical college counties), although after splitting the sample neither coefficient is individually statistically different from zero, nor is the difference between the two. Consistent with the comparison of college to consolation prize counties, I find no evidence that colleges that actively promoted technical skills induced larger increases in local invention. Appendix F.C presents several additional results for heterogeneous treatment effects across different types of colleges.[35]

---

[34] For some types of colleges, there is much more ambiguity regarding whether or not the college should be classified as practical or classical. Results are similar when using alternative classifications of practical and classical colleges; see Appendix F.C.

[35] Appendix F.D presents results of heterogeneous treatment effects based on several non-college-type county characteristics. I typically find either little heterogeneity or, if anything, larger effects when a college is established in a less-developed county. I interpret this as additional suggestive, although fairly weak, evidence that an important channel through which colleges affect local invention is by driving population and general economic development.

V Conclusion

In this paper, I document that establishing a new college causes at least 48% more patents per year in counties that receive a new college. Population increases in college counties at a similar rate to patenting. When matching patentees to college yearbooks and the U.S. census, I find that, through 1940, only about 5% of patents in a college's county come from individuals who are directly affiliated with a particular college; the majority of patents are from individuals who are affected by a new college only through indirect channels. To the extent I can observe ability in the census data, college counties do not appear to be attracting more skilled migrants than the runner-up counties. I further show that other institutions cause increases in patenting statistically indistinguishable from a college, and that colleges that focus on technical skills increase patenting indistinguishably from those that do not. Together, these suggestive results point to increasing population as the most important channel through which colleges affect local invention.

Two final caveats to this study are important. First, while alumni of a particular college played only a small role in the patenting in that college's county, this does not imply that alumni were not active inventors. Because educated individuals are highly geographically mobile, it is possible, even likely, that alumni left the counties of their alma maters to create innovations in other locations. In future work, I plan to track college alumni across time and space to determine where alumni move after they graduate and where, and if, they invent. Finally, it is important to note that promoting innovation is clearly not the only, nor even perhaps the primary, purpose of colleges and universities. Nevertheless, to the extent that policymakers wish to create inventive hubs, the results in this paper suggest that pursuing policies that drive population growth, rather than building colleges per se, can achieve the desired results.

References

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2017). The social origins of inventors. Unpublished, University of Chicago.

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

Akcigit, U., S. Baslandze, and S. Stantcheva (2016, October). Taxation and the international mobility of inventors. *American Economic Review 106*(10), 2930–2981.

Akcigit, U., J. Grigsby, and T. Nicholas (2017a, May). Immigration and the rise of American ingenuity. *American Economic Review: Papers & Proceedings 107*(5), 327–331.

Akcigit, U., J. Grigsby, and T. Nicholas (2017b). The rise of American ingenuity: innovation and inventors of the golden age. Unpublished, University of Chicago.

Almeida, P. and B. Kogut (1999, July). Localization of knowledge and the mobility of engineers in regional networks. *Management Science 45*(7), 905–917.

Andersson, R., J. M. Quigley, and M. Wilhelmsson (2004, August). University decentralization as regional policy: the Swedish experiment. *Journal of Economic Geography 4*(4), 371–388.

Andersson, R., J. M. Quigley, and M. Wilhelmsson (2009, July). Urbanization, productivity, and innovation: evidence from investment in higher education. *Journal of Urban Economics 66*(1), 2–15.

Andrews, M. J. (2019a). Comparing historical patent datasets. Unpublished, NBER.

Andrews, M. J. (2019b). Go forth and serve: geographic spillovers from colleges. Unpublished, NBER.

Andrews, M. J. (2019c). Historical appendix: How do institutions of higher education affect local invention? Evidence from the establishment of U.S. colleges. Unpublished, NBER.

Arora, A., S. Belenzon, and H. Lee (2019). Do universities substitute for corporate scientific research? Unpublished, Duke University.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2018). How well do automated linking methods perform in historical samples? Evidence from new ground truth. Unpublished, University of Michigan.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. V. Reenen (2019, May). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics 134*(2), 647–713.

Berkes, E. (2018). Comprehensive universe of U.S. patents (CUSP): data and facts. Unpublished, Ohio State University.

Bianchi, N. and M. Giorcelli (2019). Scientific education and innovation: from technical diplomas to university STEM degrees.

Blundell, R., R. Griffith, and J. Van Reenen (1995, March). Dynamic count data models of technological innovation. *The Economic Journal 105*(429), 333–344.

Bonander, C., N. Jakobsson, F. Podestà, and M. Svensson (2016, September). Universities as engines for regional growth? Using the synthetic control method to analyze the effects of research universities. *Regional Science and Urban Economics 60*, 198–207.

Bound, J., J. Groen, G. Gézdi, and S. Turner (2004, July-August). Trade in university training: cross-state variation in the production and stock of college-educated labor. *Journal of Econometrics 121*(1-2), 143–173.

Breschi, S. and F. Lissoni (2001, July). Localised knowledge spillovers vs. innovative milieux: knowledge “tacitness” reconsidered. *Papers in Regional Science 80*(3), 255–273.

Breschi, S. and F. Lissoni (2009, July). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography 9*(4), 439–468.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011, April). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Cantoni, D. and N. Yuchtman (2014, March). Medieval universities, legal institutions, and the commercial revolution. *Quarterly Journal of Economics 129*(2), 823–887.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 349–404. Amsterdam: Elsevier.

Ciccone, A. and R. E. Hall (1996, Match). Productivity and the density of economic activity. *American Economic Review 86*(1), 54–70.

Cline, P. (1983). *Mountain campus: the story of Northern Arizona University.* Flagstaff, AZ: Northland Press.

Cowan, R. and N. Zinovyeva (2013, April). University effects of regional innovation. *Research Policy 42*(3), 788–800.

Currie, J. and E. Moretti (2003, November). Mother's education and the intergenerational transmission of human capital: evidence from college openings. *Quarterly Journal of Economics 118*(4), 1495–1532.

Drucker, J. and H. Goldstein (2007, January). Assessing the regional economic development impacts of universities: a review of current approaches. *International Regional Science Review 30*(1), 20–46.

Feigenbaum, J. J. (2018, July). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *The Economic Journal 128*(612), F446–481.

Ferrie, J. P. (1996, Fall). A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 29*(4), 141–156.

Ferrier, W. W. (1930). *Origin and development of the University of California.* Berkeley, CA: The Sather Gate Book Shop.

Florida, R. (2002a, December). The economic geography of talent. *Annals of the Association of American Geographers 92*(4), 743–755.

Florida, R. (2002b). *The rise of the creative class: and how it's transforming work, leisure, community and everday life.* New York: Basic Books.

Frenette, M. (2009, June). Do universities benefit local youth? Evidence from the creation of new universities. *Economics of Education Review 28*, 318–328.

Furman, J. L. and M. J. MacGarvie (2007, August). Academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry. *Journal of Economic Behavior & Organization 63*(4), 756–776.

Geiger, L. G. (1958). *University of the northern plains: a history of the University of North Dakota, 1883-1958.* Grand Forks, ND: The University of North Dakota Press.

Glaeser, E. L. and J. D. Gottlieb (2009, December). The wealth of cities: agglomeration economics and spatial equilibrium in the United States. *Journal of Economic Literature 47*(4), 983–1028.

Goldin, C. and L. F. Katz (1999, Winter). The shaping of higher education: the formative years in the United States, 1890 to 1940. *Journal of Economic Perspectives 13*(1), 37–62.

Goldin, C. and L. F. Katz (2008). *The race between education and technology.* Cambridge, MA: The Belknap Press of the Harvard University Press.

Gordon, R. J. (2016). *The rise and fall of American growth: the U.S. standard of living since the Civil War*. Princeton, NJ: Princeton University Press.

Greenstone, M., R. Hornbeck, and E. Moretti (2010, June). Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Griliches, Z. (1990, December). Patent statistics as economic indicators: a survey. *Journal of Economic Literature 28*(4), 1661–1707.

Groen, J. A. (2004, July-August). The effect of college location on migration of college-educated labor. *Journal of Econometrics 121*(1-2), 125–142.

Hausman, N. (2017). University innovation and local economic growth. Unpublished, Hebrew University.

Helmers, C. and H. G. Overman (2017, September). My precious! The location and diffusion of scientific research: evidence from the Synchrotron Diamond Light Source. *The Economic Journal 127*(604), 2006–2040.

Holley, K. A. and M. S. Harris (2016). Universitites as anchor institutions: economic and social potential for urban development. In M. B. Paulsen (Ed.), *Higher Education: Handbook of Theory and Research*, Volume 31, pp. 393–439. New York: Springer.

Jaffe, A. B. (1989, Dec). Real effects of academic research. *American Economic Review 79*(5), 957–970.

Jung, T. and O. Ejermo (2014, July). Demographic patterns and trends in patenting: gender, age, and education of inventors. *Technological Forecasting and Social Change 86*, 110–124.

Kantor, S. and A. Whalley (2014, March). 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, April). Research proximity and productivity: long-term evidence from agriculture. *Journal of Political Economy 127*(2), 819–854.

Kerr, W. and W. Lincoln (2010, July). The supply side of innovation: H-1B visa reforms and US ethnic invention. *Journal of Labor Economics 28*(3), 473–508.

Kim, J. (2020). Towns and gowns: the impact of public universities on local innovation and entrepreneurship. Unpublished, Boston University.

Lee, J. (2019, January). The local economic impact of a large research university: evidence from UC Merced. *Economic Inquiry 57*(1), 316–332.

Leggett, M. D. (1874). *Subject-matter index of patents for invention issued by the United States Patent Office from 1790 to 1873, inclusive*, Volume 1-3. Washington, DC: Government Printing Office.

Lehnert, P., C. Pfister, and U. Backes-Gellner (2020). Employment of R&D personnel after and educational supply shock: effects of the introduction of Universities of Applied Sciences in Switzerland. Unpublished, University of Zurich.

Leten, B., P. Landoni, and B. V. Looy (2014, October). Science or graduates: how do firms benefit from the proximity of universities? *Research Policy 43*(8), 1398–1412.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014, July). Disambiguation and co-authorship networks of the U.S. Patent Inventor Database (1975-2010). *Research Policy 43*(6), 941–955.

Liu, S. (2015, May). Spillovers from universities: evidence from the land-grant program.

Journal of Urban Economics 87, 25–41.

Maloney, W. F. and F. V. Caicedo (2017). Engineering growth: innovative capacity and development in the Americas. Unpublished, IZA.

Manson, S., J. Schroeder, D. V. Riper, and S. Ruggles (2017). IPUMS national historical geographic information system: version 12.0. `http://doi.org/10.18128/D050.V12.0`.

Marco, A. C., M. Carley, S. Jackson, and A. F. Myers (2015, June). The USPTO historical patent data files: two centuries of invention. Unpublished, USPTO Economic Working Paper No. 2015-1.

Martin, D. D. (1960). *The lamp in the desert: the story of the University of Arizona*. Tucson, AZ: The University of Arizona Press.

McMath Jr., R. C., R. H. Bayor, J. E. Brittain, L. Foster, A. W. Giebelhaus, and G. M. Reed (1985). *Engineering the new south: Georgia Tech, 1885-1985*. Athens, GA: The University of Georgia Press.

Moretti, E. (2004, July-August). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics 121*(1-2), 175–212.

Moretti, E. (2012). *The new geography of jobs*. New York: Houghton Mifflin Harcourt Publishing Company.

Moretti, E. and D. J. Wilson (2014, January). State incentives for innovation, star scientists and jobs: evidence from biotech. *Journal of Urban Economics 79*, 20–38.

Moretti, E. and D. J. Wilson (2017, July). The effect of state taxes on the geographical location of top earners: evidence from star scientists. *American Economic Review 107*(7), 1858–1903.

Mowery, D. C. and B. N. Sampat (2001, August). University patents and patent policy debates in the USA, 1925-1980. *Industrial and Corporate Change 10*(3), 781–814.

Mowery, D. C. and A. A. Ziedonis (2002, March). Academic patent quality and quantity before and after the Bayh-Dole Act in the United States. *Research Policy 31*(3), 399–418.

Nagaoka, S., K. Motohashi, and A. Goto (2010). Patent statistics as an innovation indicator. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 2, pp. 1083–1127. Amsterdam: Elsevier.

O'Mara, M. P. (2005). *Cities of knowledge: Cold War science and the search for the next Silicon Valley*. Princeton, NJ: Princeton University Press.

Patrick, C. (2016, October). Identifying the local economic development effects of million dollar facilities. *Economic Inquiry 54*(4), 1737–1762.

Petralia, S., P.-A. Balland, and D. L. Rigby (2016a, August). Data descriptor: unveiling the geography of historical patents in the united states from 1836-1975. *Nature: Scientific Data 3*, 1–14. Article number: 160074.

Petralia, S., P.-A. Balland, and D. L. Rigby (2016b). Historical patent dataset. `http://dx.doi.org/10.7910/DVN/BPC15W`, accessed September 15, 2016.

Sampat, B. N. (2006, July). Patenting and US academic research in the 20th century: the world before and after Bayh-Dole. *Research Policy 35*(6), 772–789.

Sarada, M. J. Andrews, and N. L. Ziebarth (2019, October). Changes in the demographics of American inventors, 1870-1940. *Explorations in Economic History 74*.

Sumell, A., P. Stephan, and J. Adams (2008). Capturing knowledge: the location decision of new Ph.Ds working in industry. In *Science and engineering careers in the United States: an analysis of markets and employment*, pp. 257–287. Chicago: University of Chicago Press.

Toivanen, O. and L. Väänänen (2016, May). Education and invention. *Review of Economics and Statistics 98*(2), 382–396.

Wagoner, J. J. (1970). *Arizona Territory, 1863-1912: a political history.* University of Arizona Press.

Waldinger, F. (2016, December). Bombs, brains, and science: the role of human and physical capital for the production of scientific knowledge. *Review of Economics and Statistics 98*(5), 811–831.

Washburn, W. E. (1979, July). Samuel Eliot Morison, historian. *The William and Mary Quarterly 36*(3), 325–352.

Watzinger, M., L. Treber, and M. Schnitzer (2018). Universities and science-based innovation in the private sector. Unpublished, LMU Munich.

Xiong, H. and Y. Zhao (2019). Denominational competition and the market provision of human capital. Unpublished, Northwestern University.

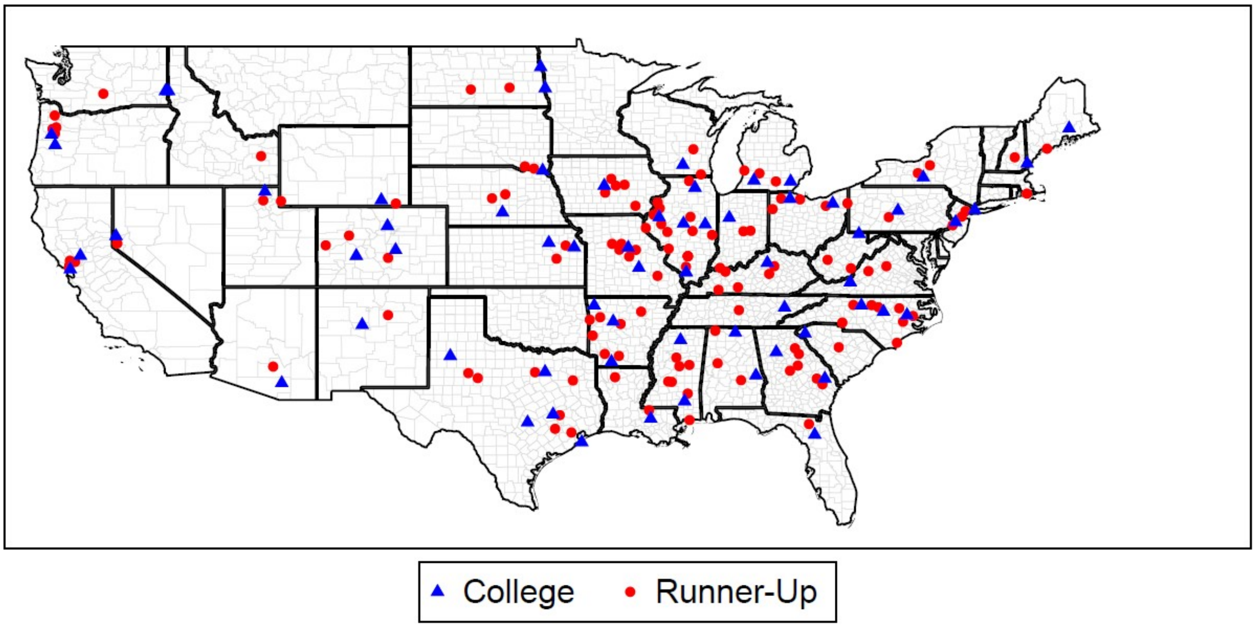
Zolas, N., N. Goldschlag, R. Jarmin, P. Stephan, J. Owen-Smith, R. F. Rosen, B. M. Allen, B. A. Weinberg, and J. I. Lane (2015, December). Wrapping it up in a person: examining employment and earnings outcomes for Ph.D. recipients. *Science 350*(6266), 1367–1371.

Zucker, L. G., M. R. Darby, and J. Armstrong (1998, January). Geographically localized knowledge: spillovers or markets? *Economic Inquiry 36*(1), 65–86.

Zucker, L. G., M. R. Darby, and M. B. Brewer (1998, March). Intellectual human capital and the birth of U.S. biotechnology enterprises. *American Economic Review 88*(1), 290–306.

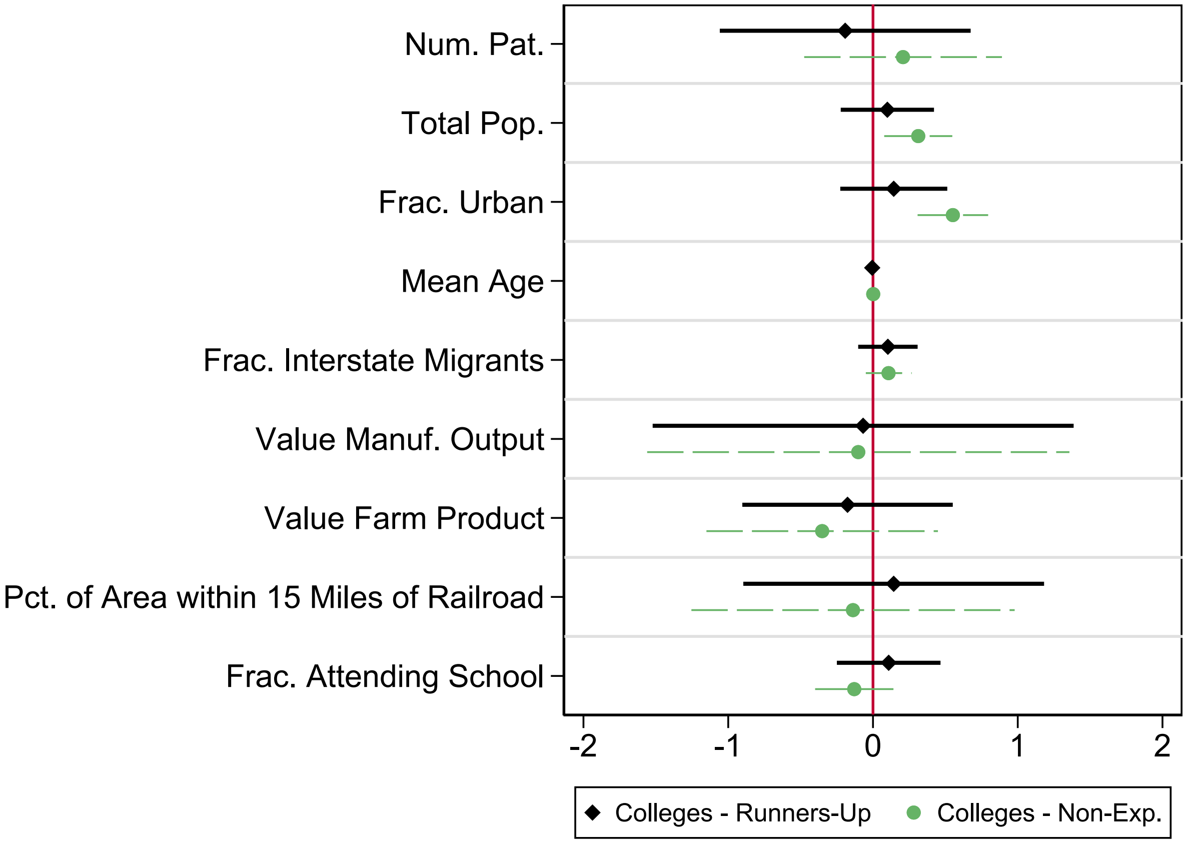
Figures

Figure 1: Map of College and Runner-Up Sites


College
Runner-Up

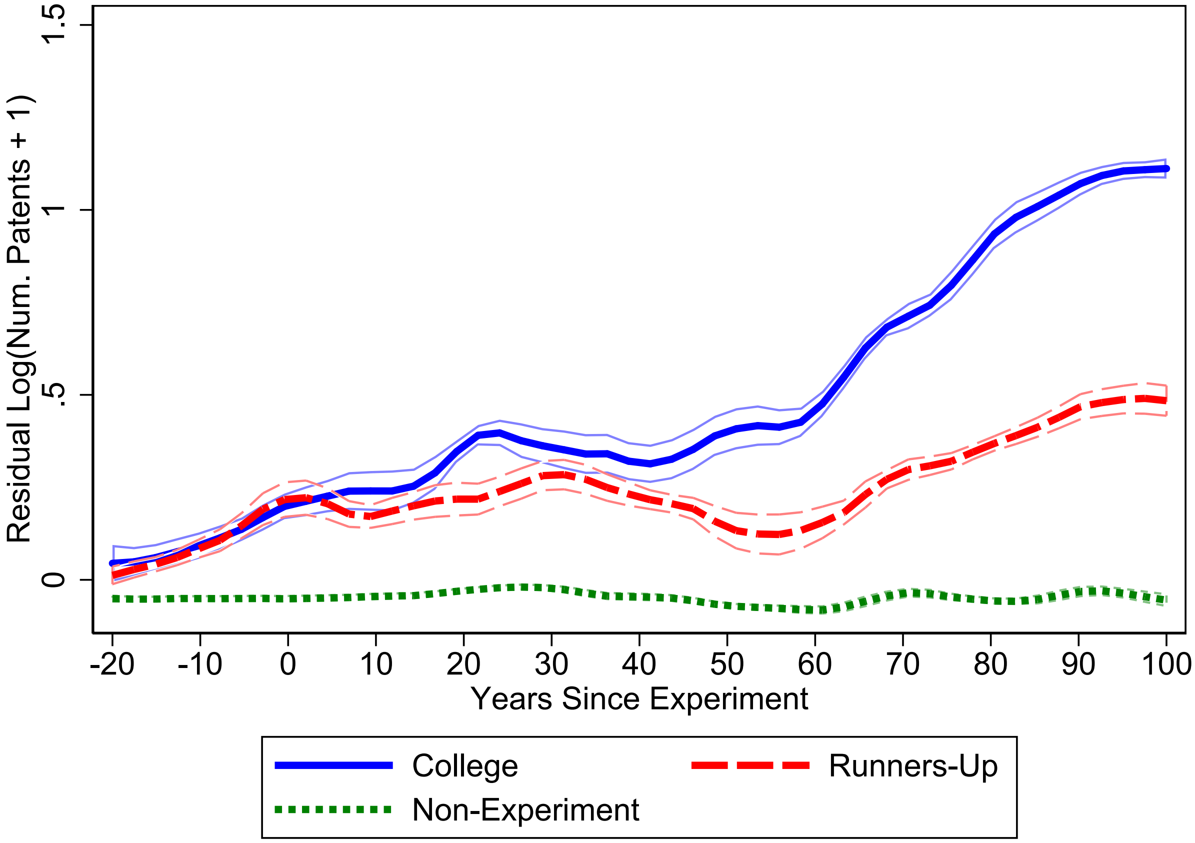
Notes: Colleges are represented by diamonds. The runner-up sites are represented by circles.

Figure 2: Balance Checks



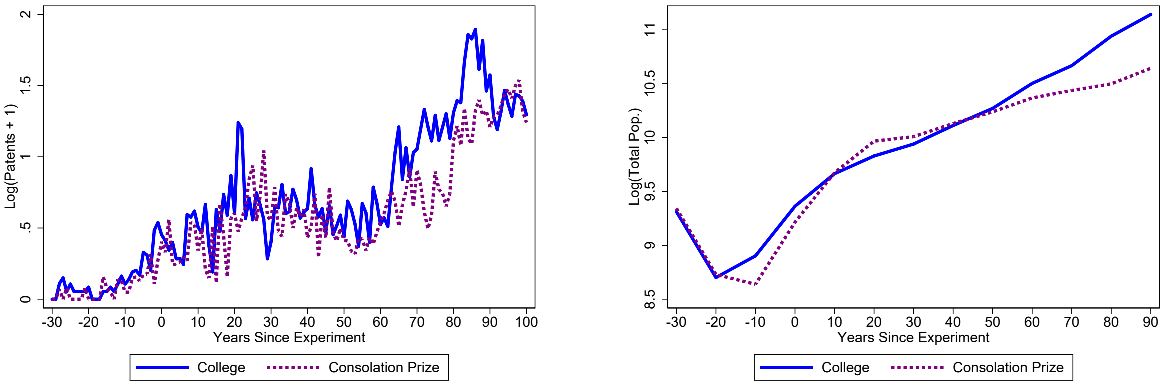
Notes: Differences in means between college counties and runner-up and non-experimental counties, scaled as a percentage of the mean in the college counties. The diamonds display the difference in the mean between the college and runner-up counties in the last U.S. census year before the college was established on a number of economic, demographic, and educational variables. The circles display the difference in the mean between the college and non-experiment counties in the last U.S. census year before the college was established. The green lines show 95% confidence intervals. The lines show 95% confidence intervals of a simple t-test of the difference in means. For readability, all values are winsorized at the 99th percentile.

Figure 3: Patenting in College, Runner-Up, and Non-Experimental Counties



Notes: Residual mean patenting in college and runner-up counties after controlling for year effects. The x-axis shows the number of years since the college experiment. The year of the college experiment is normalized to year 0. Everything left of year 0 shows pre-college means; everything to the right shows post-college means. The y-axis shows smoothed $log(Patents + 1)$. The smoothed patenting is constructed by regressing $log(Patents + 1)$ on year effects and then plotting the residuals using local mean smoothing with an Epanechnikov kernel function. The college counties are represented by the solid line. The runner-up counties are represented by the long-dashed line. The non-experimental counties are represented by the short-dashed line. Data are for high quality experiments only.

Figure 4: Patenting and Population in College and Consolation Prize Counties



(a) log(Num. Patents + 1)

(b) log(Total Pop.)

Notes: The x-axis shows the number of years since the college experiment. The year of the establishment of the new college is normalized to 0. Everything left of 0 shows pre-college results; everything to the right shows post-college results. In Panel (a), the y-axis shows log(Num. Patents + 1). In Panel (b), the y-axis shows $log(TotalPopulation)$. The college counties are represented by the solid line. The consolation prize counties are represented by the dashed line. Data are for high quality experiments only.

Tables

Table 1: Summary Statistics of College Site Selection Experiments

	N	Mean	S.D.	Min	Median	Max
# Finalist Counties	64	2.02	1.29	1.00	2.00	6.00
Distance to Finalists	127	136.09	181.08	11.92	89.05	1,443.16
Experiment Year	64	1883.73	22.33	1839.00	1884.00	1954.00
Year of First Class	59	1885.36	25.72	1795.00	1887.00	1955.00
Land Grant Colleges	64	0.45	0.50	0.00	0.00	1.00
Technical Schools	64	0.09	0.29	0.00	0.00	1.00
Normal Schools	64	0.17	0.38	0.00	0.00	1.00
HBCUs	64	0.05	0.21	0.00	0.00	1.00
Military Academies	64	0.03	0.18	0.00	0.00	1.00
Other Public Colleges	64	0.17	0.38	0.00	0.00	1.00
Other Private Colleges	64	0.03	0.18	0.00	0.00	1.00

Notes: Column 1 lists the count of experiments or counties. Column 2 lists mean values, Column 3 the standard deviation, Column 4 the minimum value, Column 5 the median value, and Column 6 the maximum value. Row 1 lists the number of runner-up counties for each experiment. Row 2 lists the distance between college and runner-up sites. Row 3 lists the experiment year. Row 4 lists the year in which students began attending the college. Row 5 lists the year when the college became racially desegregated. Row 6 lists the year the college became coeducational. Rows 7-13 list the fraction of colleges that are of each college type.

Table 2: Baseline Regression Results

	log(Pat. +1)	arcsinh(Pat.)	Alt. log(Pat.)	Poisson	log(Pat. +1)	log(Total Pop.)
College * PostCollege	0.476**	0.573**	0.277*	0.890**	0.807**	0.486*
	(0.213)	(0.259)	(0.156)	(0.436)	(0.310)	(0.255)
PostCollege	-0.056	-0.055	-0.065	6.545***	0.060	0.214*
	(0.079)	(0.092)	(0.064)	(1.532)	(0.139)	(0.124)
Zero Pat. Dummy			-0.747***			
			(0.012)			
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1840-2000
Cnty-Year Obs.	33,660	33,660	33,660	33,660	9,017	3,062
# Counties	176	176	176	176	52	176
# Experiments	64	64	64	64	64	64
Mean of Dep. Var.	0.498	0.634	0.498	1.104	0.384	9.759
Adj. R-Sqr.	0.520	0.521	0.691		0.529	0.721
Log-Likelihood				-386,186.079		

Notes: Column 1 estimates the level shift in patenting in college counties relative to runner-up counties after establishment of a new college when the dependent variable is $log(Num.Patents + 1)$. The dependent variable in Column 2 is the inverse hyperbolic sin of patents. The dependent variable in Column 3 is $log(Num.Patents)$, with values replaced with 0 if $Num.Patents = 0$ and a dummy variable for zero patents included. Column 4 presents results for a negative binomial regression. Column 5 repeats the specification in Column 1 but uses only the sample of college experiments for which yearbook data are available. The dependent variable in Column 6 is $log(TotalPopulation)$. These results use high quality experiments only. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 3: Patents by Alumni, Faculty, and Others

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	512.386	0.109	0.119	0.053	2.323
Faculty	19.013	0.004	0.001	0.000	0.474
Pre-College Others	1,159.337	0.248	0.411	0.183	3.541
Post-College Others	2,990.052	0.639	1.717	0.764	5.741

Notes: The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table 4: Difference in Income Between College and Runner-Up Counties

	log(Inc.)	log(Mig. Inc.)	log(Pat. Inc.)	log(Mig. Pat. Inc.)	Share in Top Inventor Occupations
Coll.County	-0.014	-0.004	-0.002	-0.213	-0.075
	(0.032)	(0.043)	(0.067)	(0.121)	(0.077)
Experiment FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1845-1944	1845-1944	1845-1944	1845-1944	1845-1944
Cnty-Year Obs.	2,270	1,536	287	59	2,863
# Counties	60	46	44	11	60
# Experiments	32	32	32	32	32
Mean of Dep. Var.	2.998	2.979	3.219	3.184	0.018
Adj. R-Sqr.	0.701	0.480	0.182	0.441	0.861

Notes: Column 1 estimates the difference in logged income between colleges and runner-up counties for all individuals in the county. Column 2 estimates the difference in logged income for all non-alumni, non-faculty individuals who were not present in the college or runner-up counties when the college was established. Column 3 estimates the difference in logged income for all patentees. Column 4 estimates the difference in logged income for all individuals in the Column 2 sample who are also patentees. The dependent variable in all columns is $\log(OccupationalIncome)$. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 5: Consolation Prize and College Type Results

	State Univs.	Cons. Prize	No Cons. Prize	College Types
College * PostCollege	0.528**	0.028	0.614**	
	(0.213)	(0.350)	(0.250)	
PostCollege	-0.023	-0.073	-0.008	
	(0.100)	(0.272)	(0.110)	
Practical College Interaction				0.306
				(0.198)
Classical College Interaction				0.681
				(0.522)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	19,317	4,022	15,295	33,660
# Counties	117	27	91	176
# Experiments	43	13	33	64
Mean of Dep. Var.	0.377	0.242	0.415	0.498
Adj. R-Sqr.	0.532	0.504	0.539	0.521

Notes: Column 1 estimates the baseline level shift in patenting in college counties relative to runner-up counties after establishment of a new college for the subset of experiments that could potentially involve consolation prizes. Column 2 compares college counties to only runner-up counties that receive a consolation prize. Column 3 repeats Column 1 but excludes all counties that receives a consolation prize. Column 4 shows results with separate interactions for practical colleges and classical colleges. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Online Appendix to:

How Do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges

Michael Andrews*

March 28, 2020

A More Information on the College Site Selection Experiments

Table A1 lists each high quality college site selection experiment, the county and state of the college, the runner-up counties that were considered as sites for the college, the experiment year, and the type of college established. The dates listed on this table are the date at which uncertainty over the college site location was resolved; these need not coincide with the official date of establishment for each college. In some cases, colleges have changed

---

*National Bureau of Economic Research. 1050 Massachusetts Ave., Cambridge, MA 02138. *Email*: mandrews@nber.org.

location, so the county listed need not be the current location or original location of the college. For colleges that changed location or were under consideration to change location, multiple experiments may be listed for the same college. For details on each site selection experiment, see Andrews (2019b).

Table A2 list the number of patents associated with each college site selection experiment. In the first column, the table lists the total number of patents granted in the college county and all runner-up counties over all years. Column 2 lists the total number of patents granted in the college county over all years. Column 3 lists the total number of patents granted in all runner-up counties over all years. Columns 4 and 5 list the total number of patents granted in the college and in all runner-up counties, respectively, in the years before the college is established. In spite of concerns about the sparseness of patent data for some counties, in all cases the college and runner-up counties have multiple patents throughout the sample period. In 40 of the 64 experiments, both the college and runner-up counties have at least one patent in the years before the college is established. Several of the cases in which either a college or the runner-up counties do not have a patent before the college is established were cases in which the college was established in the 1840s or 1850s and hence there were relatively few years of patent data in the pre-sample period.

A.A Comparing Sample to Non-Sample Colleges

To compare the sample to non-sample colleges, I utilize the Commissioner of Education reports from various years as described in Section I.E. For each year, these reports list the number of faculty, number of students, number of graduate students, and number of library

Table A1: List of College Site Selection Experiments

	College	County	State	Runner-Up Counties	Experiment Year	College Type
1	University of Missouri	Boone	Missouri	Saline; Cooper; Callaway; Howard; Cole	1839	Other Public
2	University of Mississippi	Lafayette	Mississippi	Montgomery; Monroe; Rankin; Winston; Attala; Harrison	1841	Other Public
3	Eastern Michigan University	Washtenaw	Michigan	Jackson	1849	Normal School
4	Pennsylvania State University	Centre	Pennsylvania	Blair	1855	Land Grant
5	The College of New Jersey	Mercer	New Jersey	Burlington; Essex; Middlesex	1855	Normal School
6	University of California Berkeley	Alameda	California	Napa; Contra Costa	1857	Land Grant
7	Iowa State University	Story	Iowa	Tama; Hardin; Polk; Jefferson; Marshall	1859	Land Grant
8	University of South Dakota	Clay	South Dakota	Bon Homme; Yankton	1862	Other Public
9	Kansas State University	Riley	Kansas	Shawnee	1863	Land Grant
10	University of Kansas	Douglas	Kansas	Lyon	1863	Other Public
11	Lincoln College (IL)	Logan	Illinois	Edgar; Macon; Warrick	1864	Other Private
12	Cornell University	Tompkins	New York	Seneca; Onondaga; Schuyler	1865	Land Grant
13	University of Maine	Penobscot	Maine	Sagadahoc	1866	Land Grant
14	University of Wisconsin	Dane	Wisconsin	Fond du Lac	1866	Land Grant
15	University of Illinois	Champaign	Illinois	Morgan; McLean	1867	Land Grant
16	West Virginia University	Monongalia	West Virginia	Greenbrier; Kanawha	1867	Land Grant
17	Oregon State University	Benton	Oregon	Marion	1868	Land Grant
18	Purdue University	Tippecanoe	Indiana	Hancock; Marion	1869	Land Grant
19	Southern Illinois University	Jackson	Illinois	Jefferson; Marion; Washington; Clinton; Perry	1869	Normal School
20	University of Tennessee	Knox	Tennessee	Rutherford	1869	Land Grant
21	Louisiana State University	East Baton Rouge	Louisiana	East Feliciana; Bienville	1870	Land Grant
22	Missouri University of Science and Technology	Phelps	Missouri	Iron	1870	Technical School
23	Texas A and M University	Brazos	Texas	Grimes; Austin	1871	Land Grant
24	University of Arkansas	Washington	Arkansas	Independence	1871	Land Grant
25	Auburn University	Lee	Alabama	Tuscaloosa; Lauderdale	1872	Land Grant
26	University of Oregon	Lane	Oregon	Washington; Polk; Linn	1872	Other Public
27	Virginia Polytechnic Institute	Montgomery	Virginia	Rockbridge; Albemarle	1872	Land Grant
28	University of Colorado	Boulder	Colorado	Fremont	1874	Other Public
29	University of Texas Austin	Travis	Texas	Smith	1881	Other Public
30	University of Texas Medical Branch	Galveston	Texas	Harris	1881	Technical School
31	North Dakota State University	Cass	North Dakota	Stutsman	1883	Land Grant
32	University of North Dakota	Grand Forks	North Dakota	Burleigh	1883	Other Public
33	University of Arizona	Pima	Arizona	Pinal	1885	Land Grant
34	University of Nevada	Washoe	Nevada	Carson City	1885	Land Grant
35	Georgia Institute of Technology	Fulton	Georgia	Clarke; Baldwin; Bibb; Greene	1886	Technical School
36	Kentucky State University	Franklin	Kentucky	Christian; Warren; Fayette; Boyle; Daviess	1886	HBCU
37	North Carolina State University	Wake	North Carolina	Mecklenburg; Lenoir	1886	Land Grant
38	University of Wyoming	Albany	Wyoming	Laramie; Uinta	1886	Land Grant
39	Utah State University	Cache	Utah	Weber	1888	Land Grant
40	Clemson University	Pickens	South Carolina	Richland	1889	Land Grant
41	University of Idaho	Latah	Idaho	Bonneville	1889	Land Grant
42	University of New Mexico	Bernalillo	New Mexico	San Miguel	1889	Other Public
43	Alabama Agricultural and Mechanical University	Madison	Alabama	Montgomery	1891	HBCU
44	University of New Hampshire	Strafford	New Hampshire	Belknap	1891	Land Grant
45	Washington State University	Whitman	Washington	Yakima	1891	Land Grant
46	North Carolina A and T University	Guilford	North Carolina	Forsyth; Durham; New Hanover; Alamance	1892	HBCU
47	Northern Illinois University	DeKalb	Illinois	Winnebago	1895	Normal School
48	Western Illinois University	McDonough	Illinois	Mercer; Hancock; Schuyler; Warren; Adams	1899	Normal School
49	University of Nebraska at Kearney	Buffalo	Nebraska	Custer; Valley	1903	Normal School
50	Western Michigan University	Kalamazoo	Michigan	Allegan; Barry	1903	Normal School
51	University of Florida	Alachua	Florida	Columbia	1905	Land Grant
52	Georgia Southern College	Bulloch	Georgia	Tattnall; Emanuel	1906	Other Public
53	University of California Davis	Yolo	California	Solano	1906	Land Grant
54	East Carolina University	Pitt	North Carolina	Beaufort; Edgecombe	1907	Technical School
55	Western State Colorado University	Gunnison	Colorado	Mesa; Garfield	1909	Normal School
56	Arkansas Tech University	Pope	Arkansas	Franklin; Conway; Sebastian	1910	Technical School
57	Bowling Green State University	Wood	Ohio	Sandusky; Van Wert; Henry	1910	Normal School
58	Kent State University	Portage	Ohio	Trumbull; Medina	1910	Normal School
59	Southern Arkansas University	Columbia	Arkansas	Polk; Ouachita; Hempstead	1910	Other Public
60	Southern Mississippi University	Forrest	Mississippi	Hinds; Jones	1910	Normal School
61	Southern Methodist University	Dallas	Texas	Tarrant	1911	Other Private
62	Texas Tech	Lubbock	Texas	Nolan; Scurry	1923	Technical School
63	US Merchant Marine Academy	Nassau	New York	Bristol	1941	Military Academy
64	US Air Force Academy	El Paso	Colorado	Walworth; Madison	1954	Military Academy

Notes: All high quality college site selection experiments in chronological order by the experiment date. Also included is the county and state of each college, the runner-up counties considered, the experiment year, and the college type of each experiment.

Table A2: Patenting in the College Site Selection Experiments

	College	Num. Pat.	Num. Pat. Coll.	Num. Pat. RunUp	Num. Pat. Coll. Pre	Num. Pat. RunUp Pre
1	University of Missouri	1509	945	564	0	0
2	University of Mississippi	330	35	295	0	0
3	Eastern Michigan University	4710	2659	2052	3	1
4	Pennsylvania State University	3388	1963	1425	2	5
5	The College of New Jersey	69224	15604	53620	5	144
6	University of California Berkeley	27566	21041	6525	0	0
7	Iowa State University	3469	560	2909	0	1
8	University of South Dakota	137	31	106	0	0
9	Kansas State University	1239	138	1101	0	0
10	University of Kansas	922	579	342	1	0
11	Lincoln College (IL)	3684	383	3300	9	17
12	Cornell University	14684	3170	11514	90	265
13	University of Maine	1055	837	218	91	22
14	University of Wisconsin	10498	8718	1780	36	69
15	University of Illinois	4343	2654	1689	9	78
16	West Virginia University	1798	60	1738	11	0
17	Oregon State University	5005	3710	1295	0	2
18	Purdue University	10749	789	9960	16	274
19	Southern Illinois University	1305	379	926	0	19
20	University of Tennessee	1368	1299	69	16	0
21	Louisiana State University	2810	2736	74	14	0
22	Missouri University of Science and Technology	542	431	111	0	2
23	Texas A and M University	160	105	55	0	0
24	University of Arkansas	305	257	48	0	2
25	Auburn University	1336	577	760	0	5
26	University of Oregon	3980	2478	1502	0	3
27	Virginia Polytechnic Institute	1967	1139	829	0	11
28	University of Colorado	13682	13395	287	0	0
29	University of Texas Austin	1353	1078	275	28	5
30	University of Texas Medical Branch	14812	1003	13809	66	34
31	North Dakota State University	1208	1009	199	0	0
32	University of North Dakota	535	155	381	0	0
33	University of Arizona	932	838	94	6	6
34	University of Nevada	3171	3119	52	9	7
35	Georgia Institute of Technology	10552	9239	1314	124	98
36	Kentucky State University	1546	63	1483	6	137
37	North Carolina State University	17520	11945	5574	28	27
38	University of Wyoming	364	58	306	7	9
39	Utah State University	2087	1206	881	1	7
40	Clemson University	2573	596	1977	3	24
41	University of Idaho	1201	190	1011	0	0
42	University of New Mexico	621	596	26	6	5
43	Alabama Agricultural and Mechanical University	4635	4335	300	30	21
44	University of New Hampshire	2229	1548	681	74	83
45	Washington State University	1389	711	678	1	2
46	North Carolina A and T University	9023	2200	6823	12	80
47	Northern Illinois University	9370	1567	7803	278	454
48	Western Illinois University	3712	394	3318	123	718
49	University of Nebraska at Kearney	381	244	137	24	21
50	Western Michigan University	5139	4342	797	515	122
51	University of Florida	357	268	89	21	8
52	Georgia Southern College	134	91	43	0	17
53	University of California Davis	2105	764	1341	36	65
54	East Carolina University	680	424	256	11	16
55	Western State Colorado University	688	112	576	19	32
56	Arkansas Tech University	471	33	439	9	65
57	Bowling Green State University	2772	1272	1500	90	504
58	Kent State University	6819	2600	4219	181	296
59	Southern Arkansas University	340	158	182	22	41
60	Southern Mississippi University	329	89	240	0	53
61	Southern Methodist University	10410	7997	2413	239	200
62	Texas Tech	426	358	68	2	5
63	US Merchant Marine Academy	20527	14217	6310	858	2178
64	US Air Force Academy	3850	664	3186	354	966

Notes: Patenting counts for each high quality college site selection experiments over all years, for each college county over all years, for each set of runner-up counties over all years, for each college county in the years before the college is established, and for each set of runner-up counties in the years before the college is established. Experiments are listed in chronological order by the experiment date.

volumes, among other variables such as tuition, for each U.S. college. It should be noted that there is no guarantee of the reliability of the Office of Education reports in each year. Indeed, for several years sample colleges are missing from the reports while the narrative histories indicate that sample colleges were in operation. This also calls into question the accuracy of the reported information in the reports. Nevertheless, these reports represent the best data available on the universe of U.S. colleges prior to the 1970s. See the Data Appendix of Goldin and Katz (1999) for a more detailed description of these data.

In each panel of Figure A1, I plot the distribution of non-experimental colleges for each variable of interest with green bars. All variables are residualized after controlling for year effects. A non-experimental college is a college that is in neither a high quality nor low quality experiment in my sample.[1] For each variable, for readability I plot the distribution over ten equal-sized bins. I then plot the ratio of the share of colleges in high quality experiments to the share of non-experimental colleges in each bin (solid line) and the share of colleges in low quality experiments to the share of non-experimental colleges (dashed line). A ratio value of one (indicated by the dark dotted line) occurs when the share of sample colleges to non-experimental colleges are equal in a given bin. Panel (a) plots the logged number of students. Both high and low quality experiment colleges have a much greater share of colleges in the larger bins, with the colleges in high quality experiments even larger than those in low quality experiments, although both ratios are close to one for the very largest student populations. Panel (b) plots the logged number of faculty and obtains the same general pattern. Panel (c) plots the logged number of graduate students. Colleges in both high and low quality experiments are also more likely than the non-experimental colleges to have a large number of graduate students, which suggests the sample colleges may be more research active. Panel (d) plots the logged number of library volumes, which proxies the colleges' role as a repository of knowledge that may be useful for driving innovation. Again, the colleges in both high and low quality experiments are over-represented in the larger bins, with colleges in low quality experiments having an even greater share of colleges in the largest two bins. While not shown, I also compare the distributions of average tuition, which is calculated by dividing each college's total tuition receipts by the number of students. In this case, the high quality experiment colleges in particular were more likely to have lower tuition than the non-experimental colleges. Kolmogorov-Smirnov tests decisively reject the null hypothesis that the distributions of high quality experiment and low quality experiment colleges are the same as the distribution of non-experimental colleges for the number (and logged number) of students, faculty, graduate students, and library volumes; these results are available upon request.

---

[1] Recall that the low quality experiments are the cases in which I can identify runner-up sites but the assignment among the runners-up is not as good as random.

In Figure A2, I conduct the same exercise but instead of comparing the sample colleges to all non-experimental colleges, I compare them to all colleges with a Carnegie Classification of R1 or R2, that is, to all institutions rated as having "high" or "very high" research activity. Data on the Carnegie Classification for each college is obtained from the IPEDS data. Across all panels, the distribution of the sample colleges is very similar to the distribution of Carnegie colleges and Kolmogorov-Smirnov tests fail to reject the null that the distributions are identical.

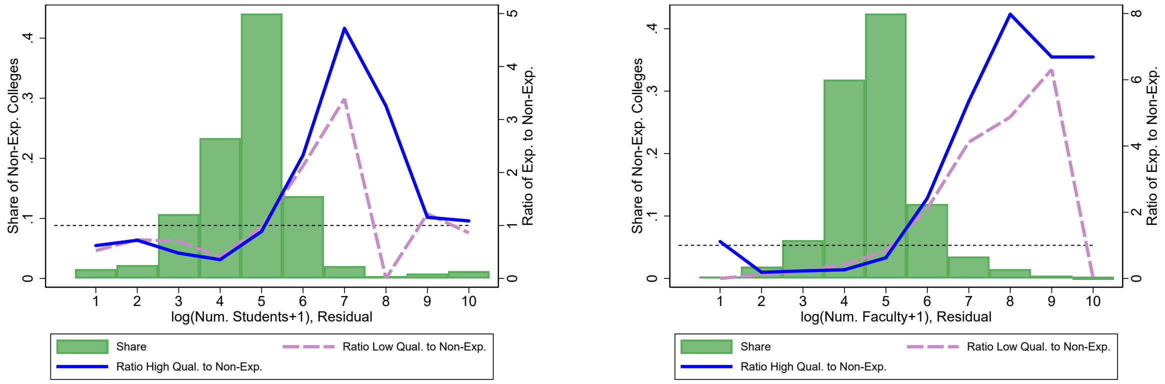
Together, these results suggest that the colleges in the sample are larger colleges than the average institution of higher education in the U.S. They are also likely to be more prominent than the average college and to be more research-focused. Indeed, on all dimensions examined the sample college appear very similar to the typical U.S. "research university" according to the Carnegie classifications. To the extent that college size, library resources, research-oriented students and faculty are important factors in determining a college's impact on the local economy, the estimates in this paper are therefore representative for large research university but likely to overstate the effects of a college relative to the "typical" college established in the U.S.

A.B Additional Balance Checks

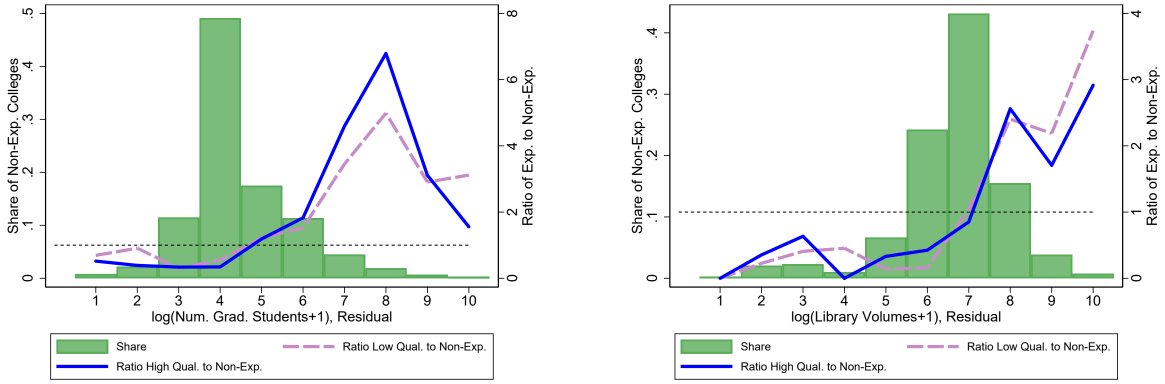
In Figure 2 in Section I.F, I compare college counties to runner-up counties along a number of observable dimensions and find that no individual dimension predicts treatment status. Here, I verify that these dimensions do not jointly predict treatment status either. Unfortunately, for several of the dimensions considered, missing data is a major concern. This is because the data come from different censuses and particular data were not necessarily collected every decade. Comparing only experiments in which data for all dimensions are available for all college and runner-up counties results in an extremely small sample size. I instead present results of joint tests with data that are available for most counties in the census year prior to the establishment of the new college.

Results of the joint tests are presented in Table A3. Column 1 estimates a linear probability model in which the dependent variable is a dummy variable taking the value of 1 when the county obtains the college and 0 otherwise. The regressors are those most likely to be correlated with both invention and the presence of a college: patenting, population,

Figure A1: Compare Colleges In Sample to All Out of the Sample Colleges



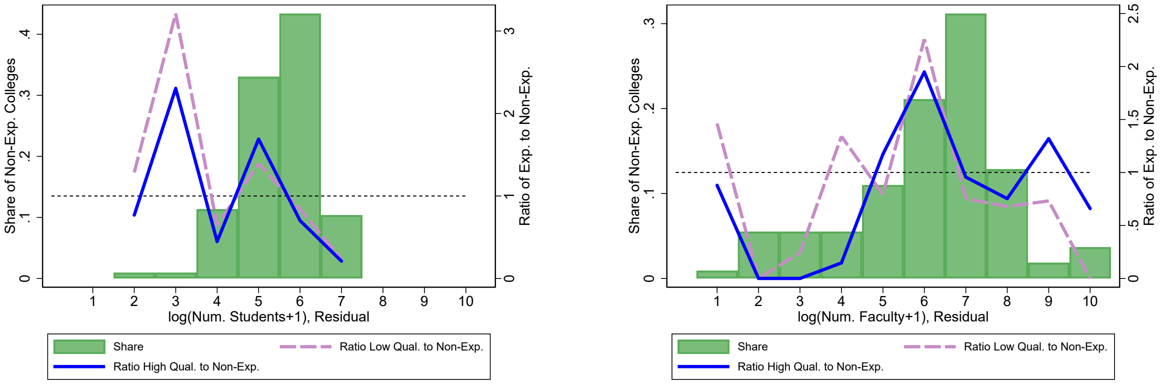
(a) log(Undergrad. Students) **(b)** log(Faculty)



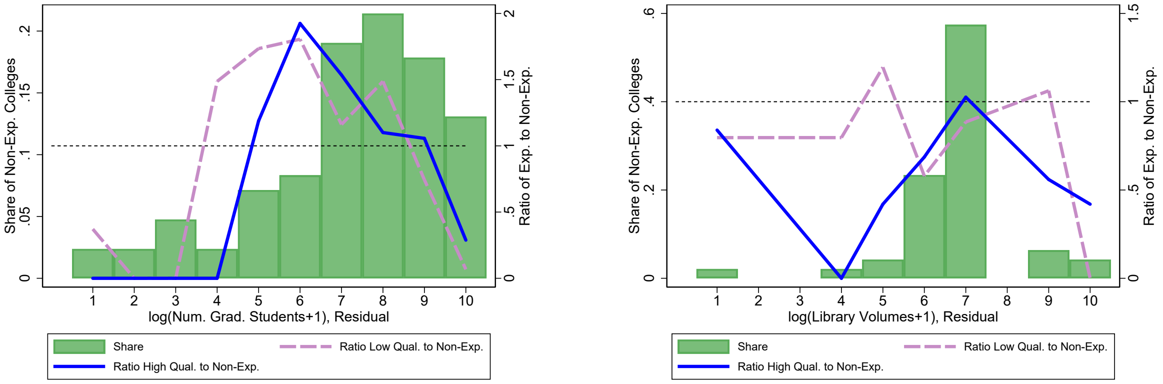
(c) log(Grad. Students) **(d)** log(Library Vols.)

Notes: The bars show the distribution of non-experimental colleges across ten equal-sized bins. The solid line plots the ratio of the share of high quality colleges to the share of non-experimental colleges in each bin. The dashed line plots the ratio of the share of low quality colleges to the share of non-experimental colleges in each bin. The dark dotted line plots a ratio of one as a reference. Panel (a) plots these results for log(Students), Panel (b) for log(Faculty), Panel (c) for log(Graduate Students), and Panel (d) for log(Library Volumes). All variables are residualized by controlling for year effects.

Figure A2: Compare Colleges In Sample to Out of the Sample Carnegie Research Institutions



(a) log(Undergrad. Students) **(b)** log(Faculty)



(c) log(Grad. Students) **(d)** log(Library Vols.)

Notes: The bars show the distribution of Carnegie R1 and R2 non-experimental colleges across ten equal-sized bins. The solid line plots the ratio of the share of high quality colleges to the share of non-experimental colleges in each bin. The dashed line plots the ratio of the share of low quality colleges to the share of non-experimental colleges in each bin. The dark dotted line plots a ratio of one as a reference. Panel (a) plots these results for log(Students), Panel (b) for log(Faculty), Panel (c) for log(Graduate Students), and Panel (d) for log(Library Volumes). All variables are residualized by controlling for year effects.

and urbanization. The F-test statistic for the joint significance of these included regressors is 1.547, which is insignificant at conventional levels. Column 2 estimates a logit model with the same regressors. A likelihood ratio χ^2-test also concludes that the regressors do not jointly predict treatment status. Columns 3 and 4 repeat Columns 1 and 2 but include all of the regressors found in Figure 2, some of which are missing for some counties and some census years, and hence these tests have much smaller sample sizes; the coefficients are again not jointly significant.[2] Results are similar with other combinations of regressors beyond those in Figure 2. Namely, I conduct joint and individual balance checks on residential segregation (see Logan and Parman (2017) for the construction of this measure), population density, manufacturing output, manufacturing establishments, manufacturing employment, manufacturing wages, farm output, farm wages, value of farms, the share of patents across patent classes, fraction of the population attending school, and fraction illiterate. I also compare logged transformations of many of these variables. In no cases are the means of these variables in college and runner-up counties statistically different from one another at the 5% level of significance. In contrast, the college and non-experimental counties are frequently statistically different from one another, with college counties appearing on average to be larger, more industrialized, more inventive, and more educated. These results are available upon request.

Figure A3 shows that not only are the levels of a number of economic and demographic variables similar in college and runner-up counties prior to establishing a new college, but the evolve similarly as well. In Panel (a), I plot logged county population for several decades both before and the establishment of a the new college in the college, runner-up, and non-experimental counties. Panel (b) plots the fraction of the county population that lives in an urban area. Panel (c) plots the logged farm output. Finally, Panel (d) plots logged manufacturing output. Plots for the other variables are similar. Confidence intervals are omitted in the figure for readability; in all cases the college and runner-up counties are statistically indistinguishable from one another.

---

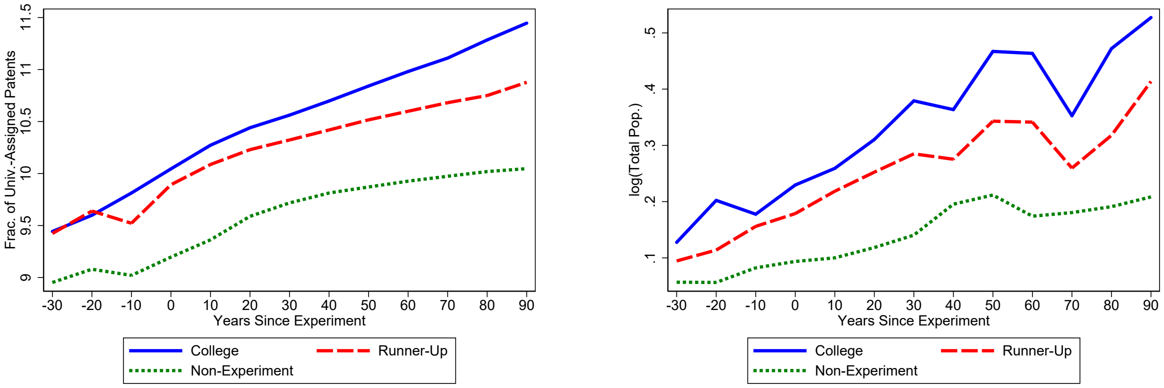
[2] Even in Columns 3 and 4, I omit data on access to railroads and on the fraction of children attending school because it is only reported in a few counties in the last census before the college is established.

Table A3: Tests for Joint Significance of Covariates Predicting Whether a County Receives a College

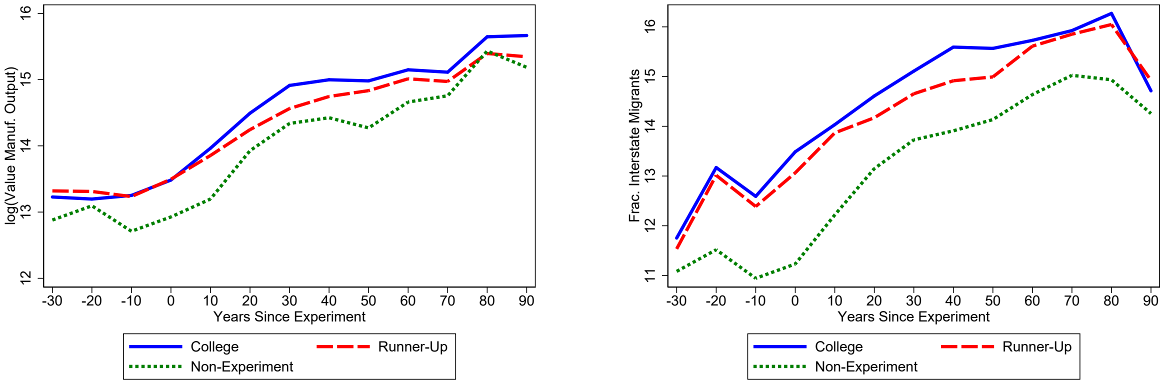
	Linear Probability	Logit	Linear Probability	Logit
log(Pat. + 1)	-0.093*	-0.450	-0.038	-0.178
	(0.051)	(0.251)	(0.091)	(0.406)
log(Total Pop.)	0.052	0.285	0.073	0.331
	(0.038)	(0.208)	(0.116)	(0.508)
Frac. Urban	0.193	0.807	-0.189	-0.990
	(0.188)	(0.834)	(0.321)	(1.445)
log(Mean Age)			-1.056	-5.458
			(1.656)	(7.464)
Frac. Interstate Migrants			0.215	1.011
			(0.236)	(1.051)
log(Value Manuf. Output)			0.058	0.306
			(0.036)	(0.202)
log(Value Farm Product)			-0.079	-0.343
			(0.073)	(0.321)
# Counties	173	173	83	83
# Experiments	60	60	59	59
Adj. R-Sqr.	0.009		-0.024	
F-Stat	1.547		0.720	
F-Test p-Value	0.204		0.655	
LR Chi-Sqr. Stat		4.874		5.534
LR-Test p-Value		0.181		0.595

Notes: Data are from the last census year before each college site selection experiment. The included covariates are those that are available for most counties in nearly every census. Columns 1 and 3 present results from linear probability models. Columns 2 and 4 present results from logit models as odds ratios. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure A3: Time Series for Demographic and Economic Variables



(a) log(Total Pop) **(b)** Frac. Urban



(c) log(Farm Output) **(d)** log(Manuf. Output)

Notes: Time series for various demographic and economic variables in each census year. The year of the college experiment is normalized to year 0. Everything left of year 0 shows pre-college means; everything to the right shows post-college means. The college counties are represented by the solid line. The runner-up counties are represented by the dashed line. The non-experimental counties are represented by the short-dashed line. In each panel, the y-axis is a demographic or economic variable. Data are for high quality experiments only.

B Constructing Patent Data

The data on patents covers the years 1836 to 2010. Patent data from before 1836 is not useful for analysis, as 1836 marked a major change in the U.S. patent system, essentially changing from a registration system to an examination system. In addition, a major fire at the U.S. Patent Office in 1836 destroyed most of the patents from the early United States, so patent records are only complete from late 1836 onward. The patent data come from four sources, with different sources available for different years. For the years 1836-1870, I use patent data collected in the Subject-Matter Index of Patents for Inventions Issued by the United States Patent Office from 1790 to 1873 (Leggett, 1874), compiled by Dr. Jim Shaw of Hutchinson, KS.[3] I use the Annual Reports of the Commissioner of Patents for the years 1870 to 1942. See Sarada, Andrews, and Ziebarth (2019) for details on cleaning, parsing, and preparing this dataset. The years 1942 to 1975 come from the HistPat dataset compiled by Petralia, Balland, and Rigby (2016a); see Petralia, Balland, and Rigby (2016b) for details on the construction of this data.[4] Finally, for the years 1975 to 2010, contemporary digitized patent data sources can be used. I utilize the data created by Li, Lai, D'Amour, Doolin, Sun, Torvik, Yu, and Fleming (2014) which contains cleaned inventor names. Because all analysis include year effects, there is no concern with the fact that different years make use of different patent data sources. Each of these datasets contains, for every granted U.S. patent, the names and residence of all inventors.[5] The fact that each patent dataset used in this paper reports the names of individual inventors is important for matching patentees to other datasets, namely college yearbook data or the U.S. population censuses. Other patent datasets that are commonly used in the literature, such as the NBER patent data and its supplements (Hall, Jaffe, and Trajtenberg, 2001), only include patents that are assigned to firms or other institutional entities and do not include the names of inventors.[6]

---

[3] See Miller (2016a) and Miller (2016b) for more information on how this dataset is compiled.

[4] I also use the HistPat data for 1874. No Annual Report could be located for that year.

[5] The Jim Shaw, Annual Reports, and Li et al. (2014) data report the town and state of each inventor; the HistPat data reports the county and state of each inventor.

To obtain additional patent-level information, I merge by patent number from the Jim Shaw, Hist Pat, and Li et al. (2014) data to other datasets that include additional patent information. In particular, I merge to the U.S. Patent and Trademark Office's Historical Patent Data Files (Marco, Carley, Jackson, and Myers, 2015), which contain information on patent classes, and the Comprehensive U.S. Patent (CUSP) Data compiled by Berkes (2018), which contain data on patent citations and patent claims. The Annual Reports do not generally have patent numbers in a usable form, so I merge to the other datasets using inventor name and town and state of residence.

For the results in this paper, I aggregate all patents to the county level. I do this for a number of reasons. First, the HistPat data records inventors' counties of residence, rather than town, and so analyzing results at a less aggregated level is impossible for this data. Second, because towns can be very small, in many cases individuals may live in one town but commute to another, even before the widespread adoption of the automobile. Aggregating to the county level thus increases the probability that a patent will be recorded in the geographic area in which the inventor actually made the invention. Moreover, individuals self report their town, with the Patent Office having no uniform way to record residences. As an example of why this is an issue, consider the example of individuals living close to Penn State University. Some may list their town as “Happy Valley,” which can refer to any of the boroughs or townships in the immediate vicinity of Penn State, while others may report “State College,” “College Township,” or one of the other adjacent townships. Aggregating to the county level avoids these issues. As the above example suggests, there is also much more variation in the names used to record particular towns. Town names are also much more likely to change over time, and new towns are incorporated and unincorporated, making it difficult to create a consistent time series of patents coming from the same geographic area. Finally, many other supplementary datasets, such as the NHGIS, are available at the county level for all years, but not at the town level.

---

[6] The listed name on the patentee is likely to be an accurate record of the individual who created the invention. Each patent is legally requird to list the name of the "first and true inventor" of a particular invention rather than, for instance, the owner of the firm in which the inventor is employed. Failure to accurately list the inventors on a patent can result in loss of patent rights, providing confidence that recorded inventor names are accurate up to transcription and character recognition errors; see Khan (2005) for more details.

Determining the county of each patent is non-trivial because each patent lists the town and state of each inventor, but not the county.[7] To match towns to their counties, I first standardize all town and county names by converting all characters to have consistent capitalization; removing all spaces, punctuation, and non-alphabetic characters; and harmonizing common abbreviations, for instance changing “SAINT” to “ST” and “FORT” to “FT”. I further manually clean some known spelling mistakes. I then obtain a list of all towns in each U.S. county in each decennial census year, compiled from the 100% censuses. I look for exact matches between town names in the patents and town names in the preceding decennial census. This means that, for instance, town names in 1883 patents are matched to town names in the 1880 decennial census. For 1890, the 100% decennial census was destroyed by fire, so I match town names to the 1990 census. The results are insensitive to matching to the closest census rather than the previous census.

---

[7] Except for the HistPat data, as described above.

For all patents granted in 1950 or later, there is no declassified 100% decennial census from the previous decade to match to. In these cases, I first attempt to match to town names in the 1940 decennial census. For the remaining towns that are unmatched, I use zip code data from `https://www.unitedstateszipcodes.org/zip-code-database/` to match to any town name that is affiliated with a current U.S. zip code; the zip code database also contains the counties in which each town resides.

Roughly 10% of town names appear in multiple counties in the same state in the same census. While this may sometimes reflect the fact that towns sit on county borders, often they occur in counties that are not adjacent to one another. When this occurs, it is impossible to know with certainty to which county the patent should belong. In these cases, I test three alternative assumptions to create county-level patent counts. Let Pat_{tsy} be the number of patents in town i that appears in multiple counties in state s and year t. Then for each county c in state s, I calculate the number of patents from the multiple-county towns as

1. $\overline{Pat_{cst}} = \sum_i Pat_{ist}$
2. $\underline{Pat_{cst}} = 0$
3. $\widehat{Pat_{cst}} = \sum_i \frac{1}{NumCounties_{ist}} Pat_{ist}$, where $NumCounties_{ist}$ is the number of counties in state s in which town i appears in year t

$\overline{Pat_{cst}}$ is an upper bound on the number of patents in each county c in state s and year t, while $\underline{Pat_{cst}}$ is a lower bound. I use the "mean" number of patents, $\widehat{Pat_{cst}}$ for all results in this paper, but the results are nearly identical when using the upper or lower bounds instead. $\widehat{Pat_{cst}}$ is the same measure constructed by the USPTO to calculate patenting by county (US Patent and Trademark Office (2000), US Patent and Trademark Office (2018)).

Errors may also occur if spelling, transcription, or OCR errors occur in town names or if the patent data use uncommon abbreviations or other slight variations of actual town names. In the baseline results presented throughout the paper, I require standardized town names in the patent data to exactly match standardized town names in the town-county correspondences. I also match towns to counties using "fuzzy" matching techniques. These are bi-gram string comparators that return a "distance" between the town-state strings in each dataset; see Andrews (2019a) for more information on the differences between the exact and fuzzy matching between towns and counties. Standardizing the town and county names eliminates most differences, and so the fuzzy matching approaches result in similar patent counts by county.

In cases when the town of a college or runner-up site changes counties over time (due either to changing county boundaries, the creation of new counties, or the combination of existing counties), I aggregate these counties over all years to their largest post-1836 historical boundaries, adopting a method similar to Atack, Jaremski, and Rousseau (2014). Results are qualitatively similar using other methods to harmonize county borders over time.

I repeat the baseline results using the HistPat or CUSP historical patent data instead of the Annual Reports, as well as using the alternative methods to match town names to counties described above. In all cases, the results are similar. These results are available upon request.

C Additional Details on Constructing the Yearbook-Patent-Census Matched Data

C.A Yearbook Data

To determine whether a particular patentee is an alumni or faculty member of a particular college, I digitize historical college yearbooks to obtain names of individuals affiliated with each college. Scanned images of a large number of college yearbooks are available on `www.ancestry.com`. After obtaining the yearbook images, I transcribe them to obtain relevant information. Table A4 lists the colleges for which yearbook data has been transcribed, including the number of yearbooks available for each college, the first and last transcribed year, and the number of transcribed records for undergraduate alumni, graduate alumni, and faculty. Due to the fact that students and faculty in the yearbooks are matched to individuals in the U.S. census and the 1940 census is the most recent that is available, no yearbooks have been transcribed for years more recent than 1940.

The type of information available and formatting of each yearbook vary enormously from college to college or even by year within the same college. This makes analysis using particular types of information difficult, as it may not be available for most years. But almost all yearbooks include the names of college seniors along with their majors. Many also include seniors' hometowns, sports teams or clubs, fraternities or sororities, or professional organizations, and often this information is available for juniors or underclassmen as well. Because I am interested in constructing a list of alumni from a particular college, I keep information only for college seniors. The assumption is that the vast majority of these

Table A4: Yearbook Data Summary Statistics

	College	Num. Yearbooks	First Yearbook	Last Yearbook	Num. Undergrads	Num. Grad. Studs.	Num. Faculty
1	Auburn University	8	1916	1940	2573	7	202
2	Clemson University	5	1915	1940	1187	0	83
3	Cornell University	45	1879	1936	25857	3313	15473
4	Georgia Institute of Technology	17	1917	1940	4309	0	1468
5	Iowa State University	30	1896	1940	8058	0	538
6	Louisiana State University	7	1927	1940	3528	713	83
7	Missouri University of Science and Technology	12	1911	1940	747	0	505
8	North Carolina A and T University	1	1939	1939	97	0	42
9	North Dakota State University	17	1908	1940	2956	0	310
10	Texas Tech	2	1937	1940	710	8	133
11	University of Arizona	9	1913	1940	1583	36	438
12	University of Colorado	27	1893	1939	5743	1	1640
13	University of Maine	32	1900	1940	4851	885	4530
14	University of Missouri	33	1898	1940	9792	574	1547
15	University of Nevada	7	1901	1940	512	0	201
16	University of New Hampshire	13	1909	1940	2673	0	2022
17	University of North Dakota	5	1906	1940	920	0	68
18	Utah State University	5	1911	1939	903	0	27
19	Virginia Polytechnic Institute	18	1898	1939	2313	50	914
20	Washington State University	12	1903	1940	4136	0	317

Notes: List of colleges for which yearbooks are transcribed. For each college, also listed is the total number of yearbooks transcribed, the earliest and the most recent transcribed yearbook, and the total number of transcribed records for undergraduate students, graduate students, and faculty.

individuals go on to become alumni in the following year; juniors will become seniors in the following yearbook, so ignoring them during their pre-graduation years saves on time and expense during the transcription process and prevents accidentally inflating the number of graduates from a particular year. Yearbooks often, although not always, also include data on each faculty member, including the faculty member's name and occasionally the highest degree obtained, position and title at the university, academic subject, alma mater, or previous academic positions held.

The yearbook data are of high quality and nearly complete for the years and schools for which yearbooks are available. To determine how complete the yearbook record is, I compare the number of seniors, faculty, and graduate students listed in the yearbooks to the same schools in the same years in the Commissioner of Education reports, described in Section I.E and A.A. Table A5 lists the mean and standard deviation of each group in the yearbooks and the Commissioner of Education reports, as well as listing the ratio of each.

Because the yearbooks and reports are only available in some years, in square brackets I list the number of instances in which a yearbook and report provided information on the same group from the same college in the same year. Obviously, such a comparison is not possible for the vast majority of yearbooks, and so all conclusions about the completeness of the yearbook data are tentative. Several features of the reports are worth noting. The reports do not list the number of college seniors for all years. For these years, I divide the number of undergraduate students by four to get the number of seniors. If college populations are growing over time, so that incoming classes are larger than the classes that came before, then this procedure will overstate the number of college seniors in the reports relative to the yearbooks. Likewise, if the reports misclassify preparatory or professional students as undergraduate students, this will also overstate the number of seniors in the reports. Additionally, as noted in Appendix A.A, it is unclear how reliable the data in the reports are; for instance, it is possible that colleges might inflate their enrollment or faculty counts to try and appear more prestigious or successful in their educational mission.

As shown in the first row, on average the yearbooks list about 66% as many seniors as the Commissioner of Education reports. The faculty, shown in Row 3, appear even more fully represented in the yearbooks relative to the reports, with the yearbooks having on average 76% as many faculty as the reports. Given the concerns with the reports data raised above, I consider these to be surprisingly high fractions and thus tentatively conclude that the yearbooks provide a fairly complete record of the college senior and faculty populations. Indeed, in Figure A4 I show that in many instances, the yearbooks record more students and faculty than do the Commissioner of Education reports.

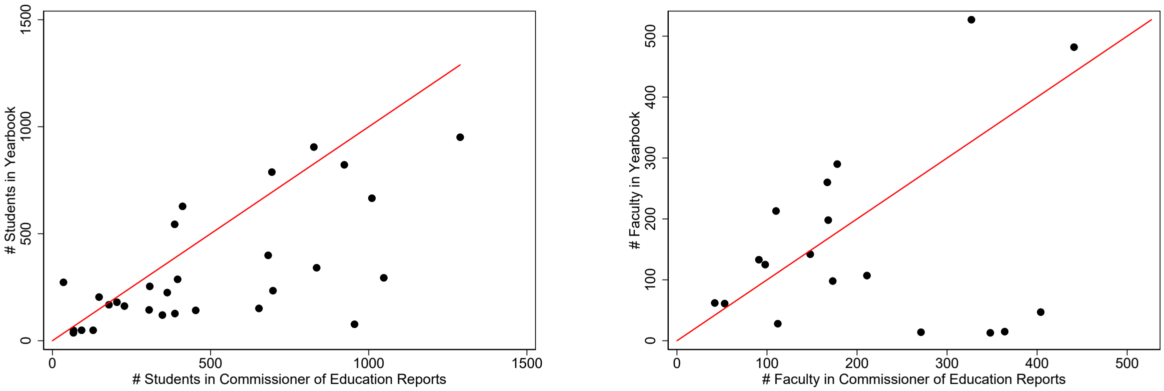
The yearbooks appear to provide a less complete picture of the graduate students, as shown in Row 2, with the yearbooks recording only 24% as many graduate students as are listed in the reports. Indeed, many of the yearbooks list no graduate students at all. Graduate students are difficult to handle for other reasons, as well. It is typically impossible to know what year graduate students are expected to graduate; yearbooks rarely list how many years a student has been at the college or how long the graduate program lasts. For instance an individual just beginning their PhD might remain a graduate student for another five years before becoming an alumnus, while professional students may be in a program for only a couple of years. This is not a concern for undergraduate seniors, because the vast majority will become alumni in the following year. Therefore, in all of the results in Section III, I ignore graduate students.

Table A5: Comparing Yearbooks to Commissioner of Education Reports

	Yearbooks	Comm. Ed. Rep.	YB / Comm.Ed.
Num. Seniors	319.62	486.82	0.657
	(274.98)	(346.57)	(49.616)
	[29.00]	[29.00]	[29.000]
Num. Grad Students	53.04	217.33	0.244
	(106.57)	(252.65)	(56.583)
	[24.00]	[24.00]	[24.000]
Num. Faculty	156.39	205.89	0.760
	(151.81)	(123.64)	(39.783)
	[18.00]	[18.00]	[18.000]

Notes: A comparison of the number of undergraduate seniors, graduate students, and faculty in the college yearbooks and the Commissioner of Education reports. Column 1 lists the number of individuals in each group in the yearbooks. Column 2 lists the number of individuals in each group in the Commissioner of Education reports. Column 3 lists the ratio of Column 1 to Column 2. Row 1 displays results for seniors, Row 2 the results for graduate students, and Row 3 the results for faculty. Standard deviations are listed in parentheses. The number of instances in which a yearbook and Commissioner of Education report both provide information on the same group from the same college in the same year is listed in square brackets.

Figure A4: Students and Faculty Counts in Yearbooks vs. Commissioner of Education Reports



(a) Seniors

(b) Faculty

Notes: Scatter plots for the number of seniors (Panel (a)) and faculty (Panel (b)) in the yearbooks and Commissioner of Education reports for all years for which both sets of data are available.

C.B Matching Patent and Yearbook Data to the Census

To determine which patentees are alumni or faculty, I merge both the patent and yearbook data to the U.S. 100% decennial population census records, transcribed by `ancestry.com` and the Minnesota Population Center and hosted by the NBER. I proceed in eight steps.

1. I prepare the census data for each census from 1850, 1860, 1870, 1880, 1900, 1920, 1930, and 1940. I restrict attention to males.[8] For each county in the census, I then link to records to the same county in the previous census, using a matching procedure that is a simplified version of Ferrie (1996) and including common names. Doing this for all censuses allows me to identify the earliest year in which a particular name appears in a particular county; I am interested in determining whether individuals first appear in a county before or after the establishment of a new college.

---

[8] I restrict attention to males for two reasons. First, women are likely to change their names between the time they show up in the yearbook data and when they patent later in life. Second, the majority of women were not a part of the labor force during the sample period, and so occupational scores, used in Section IV.A, are not informative for them.

2. I prepare counts of alumni from the yearbook data. To convert the flow of seniors listed in the yearbook of college in county i to the stock of alumni of college county i in each year T, I calculate

$$Alumni_{iT} = \sum_{t=\underline{t}}^{T} Seniors_{it},$$

   where I choose $\underline{t} = T - 60$. Under the assumption that college seniors are ≈20 years old, this means that a particular college senior can plausibly be an alumnus patentee for the next 60 years. This essentially imposes the assumption that individuals >>80 years old cannot be patentees. Such an assumption appears innocuous, as studies conclude that very few inventors are older than 80.[9]

3. I match by first name, last name, state, and county from the patent record to the census record. This creates a list of all patents in each county for which personal information about the patentees can be known. See Sarada et al. (2019) for more details on the patent-census matching procedure. I match individuals to the "closest" census. For example, for the 1900 census, I match patentees from 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, and 1904.

4. I match the lists of potential alumni and current faculty to the census, again matching on first name, last name, county, and state. I again match yearbooks to the closest census.

---

[9] For examinations of inventor ages prior to 1940, see Sarada et al. (2019) and Akcigit, Grigsby, and Nicholas (2017). Papers that document ages of more recent inventors include Jones (2009), Jung and Ejermo (2014), and Acemoglu, Akcigit, and Celik (2014).

5. I use the matched census-patent-yearbook data to determine which patentees are alumni. I calculate an alumnus patenting rate for each college county i,

$$AlumniPat.Rate_i = \frac{1}{1940 - t_0} \sum_{t=t_0}^{1940} \frac{Num.AlumniPat._{it}}{Num.Alumni_{it}}$$

   Note that both the numerator and denominator are only for those alumni and patents that I am able to match to the respective census.

6. An adjustment must be made because yearbook data are not available for all years. Without such an adjustment, the calculated stock of alumni would be too small, and if many yearbooks are missing, this omission may result in sizable undercounts of alumni patenting. To correct for this, I interpolate the number of seniors attending the college in the years in between collected yearbooks.[10] I then increase the size of the alumni stock by that number of students for each successive year,

$$\widetilde{Alumni}_{iT} = \sum_{t=\underline{t}}^{T} \widetilde{Seniors}_{it},$$

   where $\widetilde{Seniors}_{it}$ now includes the years for which the number of seniors is interpolated.

7. I calculate the share of patents belonging to alumni, faculty, and "others" based on known names from the yearbooks.[11] That is, I initially calculate the share of patents belonging to each group without using the interpolated alumni counts; I discuss this

---

[10] I use a linear interpolation for the baseline results, but other interpolation strategies yield similar or smaller alumni counts results; see Appendix E.A.

[11] I calculate the patenting *rate* for each group exactly as I do for $AlumniPat.Rate_i$, with for each group g the rate given by $Pat.Rate_{gi} = \frac{1}{1940-t_0} \sum_{t=t_0}^{1940} \frac{Num.Pat._{git}}{Pop._{git}}$, where both the numerator and denominator are for individuals matched to the census.

final adjustment in the next step. A patentee is recorded as an alumnus if there is a positive match between the individual's name from the alumni list and a name in the same county in the census and that individual is also linked to a patent. An individual is recorded as a faculty member if the individual is not recorded as an alumnus and there is a positive match between his name from the faculty list and a patent-matched name in the same county in the census. An individual is recorded as "other" if he is neither an alumnus nor a faculty member. I further split the other group into those that appear in the census in the college or runner-up county before the year in which the college was established ("pre-college others"), and those that appear in the college or runner-up county after the college is established ("post-college others"). To do this, I use the cross-census linking procedure described in the first step. The post-college others includes both those who migrate to college or runner-up counties after the college is established as well as those who are born into those counties after the college is established, and so is an imperfect proxy for in-migration.

8. Finally, I adjust the number of alumni patents and the share of patents attributed to each group to reflect the adjustments to the alumni stock. I multiply the size of the adjusted alumni stock by the calculated alumni patenting rate to get the corrected number of patents by alumni,

$$Num.\widetilde{AlumniPat}._{it} = \widetilde{Alumni}_{it} * AlumniPat.Rate_i$$

I decrease patent counts for the faculty and others by the corresponding increase in the number of alumni patents, keeping the relative sizes of the faculty and others the same. That is, I calculate

$$Num.Pat._{git} = (Num.Pat._{it} - Num.\widetilde{AlumniPat}._{it}) * \frac{Num.Pat._{git}}{\sum_g Num.Pat._{git}},$$

for groups $g \in \{Faculty, Pre-CollegeOthers, Post-CollegeOthers\}$, $Num.Pat._{git}$ are the number of patents by members of group g in college county i and year t, and $Num.Pat._{it}$ is the total number of patents in college county i and year t.[12]

C.C Match Rates

Table A6 displays patent-to-census match rates for the college and runner-up counties in the entire sample and the yearbook sample, as well as yearbook record-to-census match rates. I match 21% of the patents to the census in the full sample and 23% in the yearbook sample. These match rates are roughly double those in Sarada et al. (2019). Several possibilities exist for this discrepancy. Likely the most important factor is that I use a more liberal criteria to consider a record a match. Because I argue that the reported alumni and faculty shares are an upper bound, a more liberal matching criteria, which may include more false positive matches, is appropriate. In Appendix E.A I show the sensitivity of results to using the same match rate used in Sarada et al. (2019). Second, in the matching procedure Sarada et al. (2019) block on state and use each inventor's town name as a criteria in the matching. Instead, I block by county but do not attempt to match town names. Third, I only match to patents to males in the census, whereas Sarada et al. (2019) match to any gender. This will decrease my match rate relative to that in Sarada et al. (2019), but females account for only 4-8% of all patents over the years I study, and so this is unlikely to have much of an effect. Finally, I match to more census years than do Sarada et al. (2019) but match only select counties; the difference in sample could explain any further discrepancies.

---

[12] I impose the additional constraint that $Num.Pat._{git} \geq 0$ for all groups g. In other words, alumni in county i and year t cannot have more patents than there were total in county i and year t, even if the adjusted alumni stock and average patenting rate would suggest this to be the case.

Only about 4% of the yearbook records match to the census. Because I use the same information and matching criteria to match the yearbook records as I do to match the patent records, and because the yearbook data are likely cleaner with fewer transcription errors, I interpret this as evidence that most of the students listed in the yearbooks are likely to out-migrate after they graduate. The fact that some of the colleges are coeducational during the yearbook years and I only attempt to match to males in the census may also depress the yearbook-to-census match rate.

Table A6: Match Rates

	Match Rates to Census
Patents	0.210
Patents (Yearbook Colleges)	0.231
Yearbooks	0.037

Notes: Row 1 displays the match rate from the patent records to the census records for all college and runner-up counties. Row 2 displays the match rate from the patent records to the census records for college and runner-up counties in the yearbook sample. Row 3 displays the match rate from all yearbook records to the census records.

C.D Details on the Yearbook Sample

As described above, the yearbook sample was selected with the intention to be representative, but subject to the constraint that yearbooks were not available for all colleges. To further explore the representativeness of the yearbook sample, in Table A7, I repeat the baseline regressions from Table 2 but using only the 20 colleges for which yearbook data are available. The coefficients are qualitatively similar, although larger in magnitude, to those in the baseline sample.

Table A7: Baseline Regression Results with the Yearbook Colleges Sample

	log(Pat. +1)	arcsinh(Pat.)	Alt. log(Pat.)	Poisson	log(Total Pop.)
College * PostCollege	0.807**	1.013**	0.531**	2.182*	0.318
	(0.310)	(0.408)	(0.220)	(1.298)	(0.650)
PostCollege	0.060	0.088	-0.015	9.448***	1.383**
	(0.139)	(0.174)	(0.115)	(3.646)	(0.550)
Zero Pat. Dummy			-0.702***		
			(0.022)		
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1840-2000
Cnty-Year Obs.	9,017	9,017	9,017	9,017	800
# Counties	52	52	52	52	52
# Experiments	18	18	18	18	18
Mean of Dep. Var.	0.384	0.486	0.384	0.872	9.348
Adj. R-Sqr.	0.529	0.527	0.690		0.672
Log-Likelihood				-70,573.173	

Notes: Column 1 estimates the level shift in patenting in college counties relative to runner-up counties after establishment of a new college when the dependent variable is $log(Num.Patents + 1)$. The dependent variable in Column 2 is the inverse hyperbolic sin of patents. The dependent variable in Column 3 is $log(Num.Patents)$, with values replaced with 0 if $Num.Patents = 0$ and a dummy variable for zero patents included. Column 4 presents results estimating a Poisson model. Results are for college counties for which yearbook data is available. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D Robustness Checks and Extensions for Baseline Results

D.A Dynamics

One concern, highlighted by the fact that the college counties tend to out-patent the runner-up counties the most many decades after the establishment of a new college, is that the baseline results are largely driven by relatively recent changes in the patent law that differentially encourage patenting in college counties relative to the runners-up. Indeed, just such a change occurred in 1980 with the passage of the Bayh-Dole Act. Prior to the passage of the Bayh-Dole Act, any patent obtained while the inventors were funded by a federal grant or research contract had to be assigned to the U.S. federal government, disincentivizing university patenting. The Act instead gave ownership of these patent rights to the inventors or their institutions, including universities. As Sampat (2006) and Hausman (2017) show, this led to a dramatic increase in patents assigned to universities.

In this section, I explore the dynamic effects of establishing a new college in more detail. In particular, I seek to understand the extent to which the observed result is purely the result of recent policy changes that give colleges a more prominent role in the U.S. innovation system, and which are effects of colleges that manifested themselves relatively quickly, prior to the second half of the twentieth century.

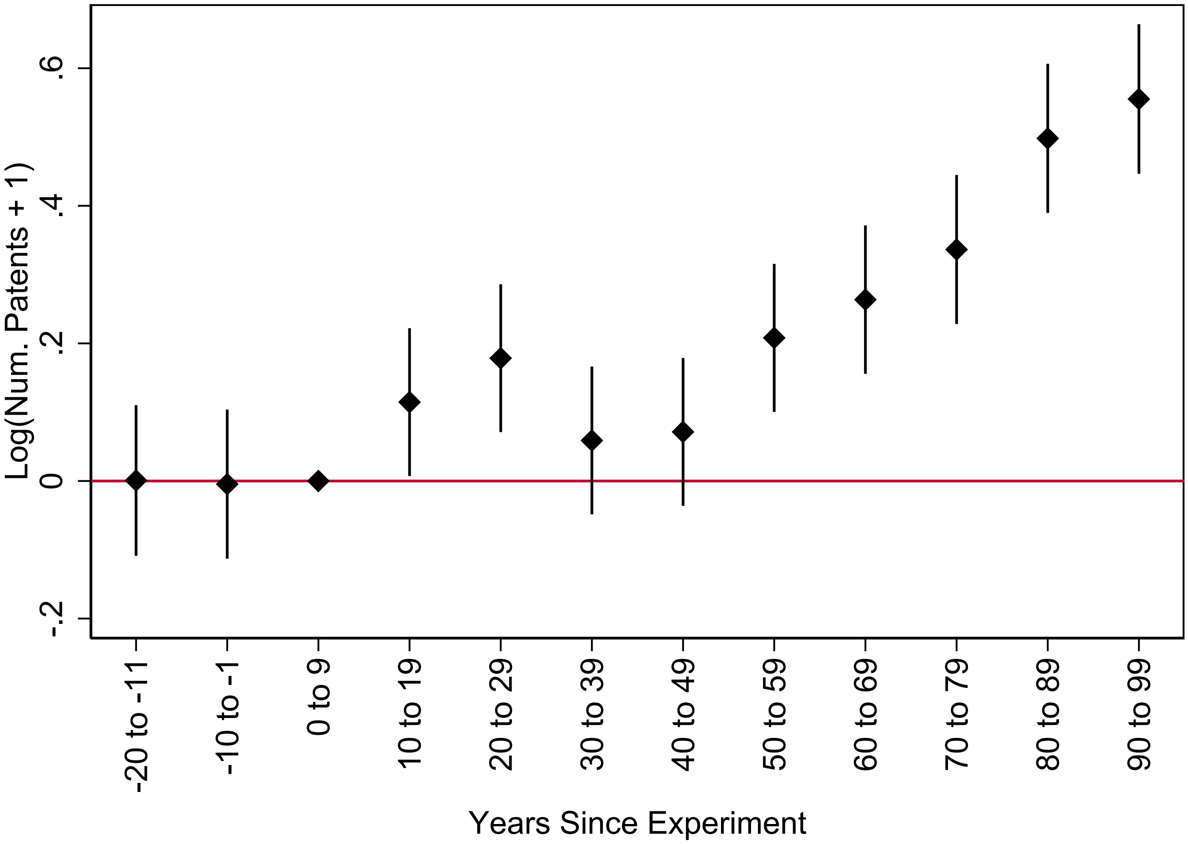
In Figure A5 I interact the effect of a college by ranges of years. More precisely, I estimate

$$\log(NumPat_{it}+1) = \sum_{\tau \in T} [\delta_{1\tau} College_i * TimeBin_\tau + \delta_{2\tau} TimeBin_\tau]$$
$$+ County_i + Year_t + \epsilon_{it},$$

where $\tau \in T$ represent "bins of years" (i.e., 0-10 years after college j is established, 10-20 years after the college is established, etc.) and $TimeBin_\tau$ is an indicator variable that is equal to one if $t \in \tau$ and 0 otherwise. Each plotted coefficient represents $\delta_{1\tau}$ in the respective range of years τ since the college establishment. The first decade in which the college is established is the omitted category. Results are nearly identical using different groupings of years. There is no significant difference between the college and runner-up counties in the decades before the establishment of each college and the estimated coefficients are close to zero in magnitude, suggesting no differential pre-trends. The interactions by decade become positive and significant after the first decade, although not all coefficients are significant at the 5% level. Confirming the intuition in Figure 3, the difference between the college and runner-up counties is largest after several decades.

In Sections III.A and E.C, I confront the possibility that observed results on patenting by alumni and faculty are due to the fact that changing policies have altered the role of colleges in the U.S. innovation system, causing patenting by these groups to explode. Instead, I present suggestive evidence that recent decades present simply a continuation of trends documented in the pre-1980 decades. To drive this point home, I repeat the baseline estimates from Table 2 but excludes all data after 1980. The estimated coefficients are only slightly smaller in magnitude. Imposing earlier cutoffs substantially reduces the precision of the estimates, although in all cases results are qualitatively similar but smaller in magnitude. For example, when excluding all data after 1940, the last census for which individual-level data is available, the estimated coefficients are roughly half the size of these results but not statistically significant at conventional levels. These results are available upon request.

Figure A5: Dynamics of Treatment Effect



Notes: Estimated coefficient of the level shift in patenting in college counties relative to runner-up counties after establishment of a new college with a separate interaction term estimated for each time bin, along with 95% confidence bands. Time bins are are dummy variables that are equal to one for college counties in every ten year period before and after the establishment of the new college.

D.B Runner-Up Counties that Receive a College at a Later Date

A related issue is how to interpret the treatment effects in light of follow-on investment that may occur. In particular, the runner-up counties may eventually receive institutions of higher education of their own.[13] After all, each runner-up site was at one point considered a nearly ideal locations for a college, so it makes sense that if there were plans to establish an additional college in the region at a later date, the runner-up counties would once again be prime candidates. While it is possible to manually check for these occurrences for large, prominent institutions, and then simply exclude all years after the later college is established in a runner-up site, the U.S. is unique in having a large number of small institutions, many of which changed names or locations and started informally, making it extremely difficult to determine the "start date" for many of these schools without a deep exploration of the narrative history of each institution. Instead, I take the more extreme step of removing from the sample any runner-up county that had a college in 2010 according to the Integrated Postsecondary Education Data System (IPEDS).[14]

Between 60% and 76% of the runner-up counties have a college in the IPEDS data, depending on what I consider to be a college. The issue of runner-up counties receiving post-treatment colleges may therefore plausibly lead to substantial underestimates of the local effect of establishing a college.

To get a sense of the extent this issue can effect estimated magnitudes, in Column 1 of Table A8, I exclude all runner-up counties that have a "traditional" college or university, defined as all institutions of higher education except for trade schools, professional schools, for-profit colleges, and community colleges. In Column 2, I also exclude runner-up counties that have a professional school, such as a specialized seminary or medical school. In Column 3 I further exclude all trade schools (e.g., cosmetology schools) and for-profit colleges (e.g., University of Phoenix campuses) that are in the IPEDS data. Finally, in Column 4 I additionally exclude all community colleges. The prevalence of these various types of institutions can be seen by the declining sample size in each column. In all cases, the results are larger than the baseline estimates, although qualitatively similar, and the magnitude increases as more institutions are excluded, until the exclusion of community colleges. This is consistent with the results in Section IV.B, which finds that excluding runner-up counties that get other types of institutions increases the estimated effect of establishing a new college. Care must be taken in attributing this interpretation to these results, as it may also be driven by heterogeneity in the types of colleges that remain in the sample after excluding runner-up counties that eventually get a college of their own. For instance, establishing a large and prominent college may decrease the need for another college in the same region at a later date.

---

[13] While this may affect the interpretation of the magnitude of the baseline results, note that all of the results in Section III observe the identities of patentees within a college county and therefore do not depend on the follow-on investment, or the lack thereof, in the runner-up counties.

[14] See `https://nces.ed.gov/ipeds/`.

In some cases, a particular county may be under consideration to receive multiple colleges *that are in my site selection experiment sample.* I exclude these counties from the results in the body of the paper. In Column 5, I include these counties. I estimate the following specification:

$$\begin{aligned} PatentMeasure_{ijt} =& \delta_1 College_{ij} * PostCollege_{jt} + \delta_2 PostCollege_{jt} \\ & + County_i + Experiment_j + County_i * Experiment_j \\ & + Year_t + \epsilon_{ijt}. \end{aligned} \tag{1}$$

The difference between this specification and the baseline specification in Equation (1) is the inclusion of county-by-experiment fixed effects to account for the fact that the same county can now appear in multiple site selection experiments. While it is less intuitive to interpret the variation in this specification, the results are qualitatively similar to, although a bit smaller than, the baseline results. This is not surprising given that there are very few cases in which the same runner-up county appears in multiple experiments.

Table A8: Runner-Up Counties with Colleges

	No Traditional Colleges	No Traditional/Professional Colleges	No Non-Community Colleges	No Colleges of Any Type	Counties in Multiple Experiments
College * PostCollege	0.722***	0.762***	0.777***	0.664**	0.309**
	(0.235)	(0.240)	(0.286)	(0.277)	(0.132)
PostCollege	-0.167**	-0.176**	-0.132	-0.107	-0.007
	(0.073)	(0.074)	(0.094)	(0.095)	(0.065)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
County-Experiment FE	No	No	No	No	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	12,883	12,533	10,608	8,532	39,784
# Counties	71	69	59	47	180
# Experiments	27	27	24	20	72
Mean of Dep. Var.	0.390	0.382	0.363	0.415	0.480
Adj. R-Sqr.	0.453	0.457	0.459	0.465	0.506

Notes: Column 1 excludes all runner-up counties with a traditional college. Column 2 also excludes all runner-up counties with a professional school. Column 3 additionally excludes all runner-up counties with a trade school or for-profit college. Column 4 additionally excludes all runner-up counties with a community college. Column 5 includes runner-up counties that appear in multiple college site selection experiments. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D.C Additional Specifications

In this section, I estimate several additional regression specifications to demonstrate that the baseline results described in Section II are robust. Results are presented in Table A9.

In Column 1, I estimate the following modification of the baseline regression:

$$\begin{aligned} PatentMeasure_{it} =& \delta_1 College_i * PostCollege_{it} + \\ & + County_i + Year_t + Experiment_j + Experiment_j * Year_t + \epsilon_{ijt}. \end{aligned} \tag{2}$$

Here, the $PostCollege_{it}$ term is not needed as it is colinear with the experiment-by-year fixed effect, $Experiment_j * Year_t$. Establishing a new college causes 36% more patents per year in the college counties than in the runner-up counties, only slightly smaller than the baseline results in Column 1 of Table 2. Column 2 examines the extensive margin: do counties have a higher probability of obtaining at least one patent per year after receiving a new college. In this linear probability model, I find that establishing a new college makes a county 26% more likely to have at least one patent in a given year. Column 3 uses an alternative construction of logged patenting, $\log(Num.Patents + 0.0001)$. These results are much larger than the baseline estimate. This is not surprising in light of the results in Column 2, since this specification penalizes having zero patents more heavily than the baseline specification that uses $\log(Num.Patents + 1)$ as the dependent variable. Column 4 displays the results using the number of patents as the dependent variable in a simple linear specification. To convert this estimate to a proportional increase, I divide the estimated increase in the number of patents by the average number of patents in 1880, around the year that the median college is established. The estimated percentage increase is large (almost 500% more patents per year in the college counties relative to the runners-up, using 1880 as the baseline year) but in line with the results using a negative binomial model in Table 2. Column 5 presents results from a fixed effects negative binomial specification, again using winsorized data and expressed as incident rate ratios. These results are nearly identical to the Poisson results presented in Table 2. In sum, while the exact magnitude varies, all specifications tell the same story: establishing a new college causes a sizable increase in local patenting.

Table A9: Additional Regression Specifications

	log(Patents + 1)	Any Patents	log(Pat. + 0.0001)	Num. Patents	Neg. Binomial
College * PostCollege	0.355***	0.261***	2.848*	5.880**	0.890**
	(0.117)	(0.083)	(1.609)	(2.563)	(0.436)
PostCollege		0.007	-0.001	-2.421	6.545***
		(0.068)	(0.313)	(1.533)	(1.532)
County FE	Yes	Yes	Yes	Yes	Yes
Experiment FE	Yes	No	No	No	No
County-Experiment FE	Yes	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes
Experiment-Year FE	Yes	No	No	No	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	33,617	33,773	33,660	33,660	33,660
# Counties	176	176	176	176	176
# Experiments	64	64	64	64	64
Mean of Dep. Var.	0.498	0.425	-5.084	1.104	1.104
Adj. R-Sqr.	0.670	0.389	0.453	0.337	0.202
Log-Likelihood	-30,690.407	-16,011.898	-94,455.024	-144,125.458	

Notes: Column 1 includes experiment-by-year fixed effects when the dependent variable is $log(Num.Patents + 1)$. Column 2 estimates a linear probability model where the dependent variable is an indicator equal to one if a county has at least one patent in a given year and zero otherwise. The dependent variable in Column 3 is $log(Num.Patents + 0.0001)$. The dependent variable in column 4 is the number of patents. Column 5 estimates a negative binomial model. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D.D Preexisting Colleges

It is important to note that there may be a distinction between establishing an additional college in a county and establishing *the first* college in a county. In the baseline results, I consider the establishment of any college for which I can identify high quality runner-up counties, independent of the presence or absence of previously established colleges in either the college or runner-up counties. In practice, the focal colleges I study were often the first colleges built in an area, particularly for western states. In cases where previous colleges existed, they were typically extremely small, with tenuous survival prospects, relative to the experimental college in my sample. Nevertheless, a college's effect on a local area may systematically differ depending on whether or not a preexisting college was present. I systematically investigate these issues in Table A10.

For the results in Column 1, I return to the narrative college histories and exclude any cases in which the presence of a preexisting college was mentioned as a factor in the college site selection decision. For instance, in the cases of several land grant colleges such as Virginia Tech, the state decided to allocate its land grant status, and enlarged state and federal support and visibility that went with it, to one of several existing public universities (Kinnear, 1972; Wallenstein, 1997).[15] When excluding these cases, establishing a college causes about 56% more patents per year in the college county relative to the runners-up, relative to a baseline increase of 48%.

In Column 2, I examine only the cases in which the presence of a preexisting college was mentioned as a factor in the college site selection decision.[16] The estimate is nearly identical to that in Column 1, with college counties producing 55% more patents per year relative to the runner-up counties.

---

[15] When a preexisting college was mentioned as a factor in a runner-up county, I omit the runner-up county. When a preexisting college was mentioned as a factor in the college county, I omit the entire experiment, dropping the college county and all runner-up counties.

The narrative histories may fail to mention all preexisting colleges in the college and runner-up counties, and even if these preexisting institutions did not affect the site selection decision, they still may have systematically influenced the new college's effect on the local economy. To account for this, in Column 3 I turn to the Commissioner of Education reports (discussed in Section I.E and A.A) and exclude any cases in which the reports list the presence of colleges in the college county in the years before the focal college is established. In addition to the concerns about the accuracy and completeness of the Commissioner of Education reports raised above, the reports are not available in the years before college establishment for all of the colleges in the sample. Nevertheless, when excluding these cases, I find results that are similar to the baseline results.

Finally, in Column 4 I exclude all experiments in which the Commissioner of Education reports list the presence of colleges in the years before the focal college is established in any of the experiment counties, not just in the college county. The result is similar to, although a bit smaller than, the baseline estimate.

In short, across all specifications, it does not appear that the presence or absence of preexisting colleges substantially alters the interpretation of the results, although the coefficients are perhaps a bit larger in cases when no college has been previously established.

---

[16] More specifically, I keep all counties in any experiment for which a preexisting college was mentioned as a factor for either the college county or any of the runner-up counties.

Table A10: Experiments with and without Preexisting Colleges

	Previous College Not Factor in Decision	Previous College Factor in Decision	No Previous Colleges in Treatment County	No Previous Colleges
College * PostCollege	0.561**	0.545**	0.445**	0.343**
	(0.283)	(0.246)	(0.216)	(0.170)
PostCollege	-0.036	0.272*	-0.038	0.038
	(0.096)	(0.160)	(0.083)	(0.094)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	26,835	10,629	31,757	26,507
# Counties	147	57	165	142
# Experiments	52	19	61	52
Mean of Dep. Var.	0.502	0.551	0.489	0.489
Adj. R-Sqr.	0.534	0.527	0.524	0.555

Notes: Column 1 excludes all runner-up counties with a traditional college. Column 2 also excludes all runner-up counties with a professional school. Column 3 additionally excludes all runner-up counties with a trade school or for-profit college. Column 4 additionally excludes all runner-up counties with a community college. Column 5 includes runner-up counties that appear in multiple college site selection experiments. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D.E Alternative Geographic Boundaries

Clearly, colleges can affect invention across county borders as well. If a college is very close to a county border, or if a county is very small, then counties may not be the best geographical unit at which to examine the results. In Column 1 of Table A11 I present results at the commuting zone level, dropping any runner-up observations that take place in the same commuting zone as the college. I use commuting zone definitions providing by the U.S. Department of Agriculture Economic Research Service for the year 1980, the earliest year for which commuting zones are defined (`https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/`). Results are nearly identical using commuting zones defined for 1990 or 2000. The estimated treatment effect is similar to the baseline result, with commuting zones that receive a new college having about 64% more patents per year relative to their runner-up commuting zones after the college is established.

D.F Alternative Samples of Colleges

I conduct a number of additional robustness checks as well. To further show that the results are not driven by the subjective classification of some experiments as either high or low quality, I re-estimate the baseline regression excluding each high quality experiment, one at a time, and re-estimating the baseline regression. I also reclassify each low quality experiment as high quality, one at a time, and re-estimate the baseline regression. In all cases, the estimated coefficient is very similar to the baseline result and statistical significance is unchanged. A related concern is that the results are driven primarily by large cities, as in the example of Georgia Tech mentioned in the Introduction. It may be a stretch to believe that all of the differences between Atlanta and Macon that occurred after the establishment of Georgia Tech were due to its creation (although Georgia Tech was likely the cause of some follow-on investment). To verify that these largest cities are not driving the results, I omit data from counties with large populations and find that the results are largely unchanged. An additional concern is that different types of college experiments may be systematically different from one another. While each experiment is unique, they tend to fall into groups in which the colleges were assigned with different general methods. It would be suspicious if one method of "random" assignment gave systematically different results from other such methods. I test this by grouping experiments by the method in which the college was assigned and then verifying that the estimated coefficients are similar across different groups. All of these results are available upon request.

D.G A Placebo Test

I next conduct a placebo test to determine whether patenting changes differentially in college and runner-up counties in the years leading up to the college site selection experiment. I drop all data for the years after and including the year in which the college was established; all the remaining data is for the pre-trend. I then artificially designate the halfway point between the first year of observations and the last pre-experiment year as the "experiment year" and re-run the baseline regressions. Results are presented in Columns 1 and 2 of Table A11, with Column 1 showing the effects on logged patenting and Column 2 showing the effects on logged county population. If the college counties are up-and-coming places, then they should be growing faster then the runner-up counties in the years before the original college site selection experiment, both in terms of the number of inventions and the size of the population, and the estimated coefficient ($College*$"$PostCollege$") should be significantly positive. Instead, neither of the coefficients are statistically different from zero and, while slightly positive, are close to zero in magnitude relative to their counterparts in Table 2. I take this as further evidence that the college site selection experiment is valid. Results are very similar if I instead designate random pre-college years as the placebo "treatment" year.

D.H Patent Classes

Establishing a new college may alter the composition of patented technologies in addition to changing the total number of patents. To get a sense of patent technology type, in Table A12 I use the patent classes assigned to historical patents by Marco et al. (2015) to examine how

Table A11: Placebo Test and Results at Other Geographic Levels

	Commuting Zones	log(Pat. + 1)	log(Pop.)
College * PostCollege	0.639**	0.078	0.150
	(0.262)	(0.081)	(0.142)
PostCollege	-0.137	-0.074*	-0.102
	(0.087)	(0.041)	(0.070)
County FE	No	Yes	Yes
Commuting Zone FE	Yes	No	No
Year FE	Yes	Yes	Yes
Year Range	1836-2010	1836-1953	1840-1950
Cnty-Year Obs.	25,460	8,784	691
# Counties	138	192	175
# Experiments	71	64	64
Mean of Dep. Var.	0.476	0.410	9.535
Adj. R-Sqr.	0.538	0.216	0.553

Notes: In Column 1, I present baseline regression results estimated at the commuting zone level rather than the county level. In Columns 2 and 3, the baseline regression results are reproduced with all post-experiment data dropped. The experiment year is set to halfway between the initial year of patent data and the year prior to the original college site selection experiment. In Column 2 the dependent variable is logged patenting, while in Column 3 it is logged population. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

patenting across all classes changes after establishing a new college.[17] In Column 1, I include controls for the share of patents in each county that belong to each of the NBER patent classes (patents with missing classes is the omitted category). The differences-in-differences estimate is a bit more than half the size of the baseline estimate and is no longer statistically significant at conventional levels.

In Column 2, I repeat the baseline estimate at the patent class-by-county-by-year level. That is, I estimate:

$$\begin{aligned} PatentMeasure_{ijct} =& \delta_1 College_i * PostCollege_{it} + \delta_2 PostCollege_{it} \\ & + County_i + Class_c + Year_t + Class_c * Year_t + \epsilon_{ijct}, \end{aligned} \tag{3}$$

for patent classes c. This specification thus includes patent class and patent class-by-year fixed effects, flexibly picking up the fact that certain types of technology may be more or less prevalent at different points in time. The coefficient is similar to that in Column 1.

The results in Columns 1 and 2 suggest that there is some shifting in the composition of types of inventions patented in college counties after a new college is established. Are college counties becoming increasingly specialized in a few narrow technology areas that happen to be especially patent-prone? This does not appear to be the case. To see this, I construct a Herfindahl-Hirschman index of patent concentration:

$$Pat.Concent_{it} = \sum_{c \in C_{it}} \Big(\frac{Num.Pat_c}{\sum_{k \in C_{it}} Num.Pat_k} \Big)^2 \tag{4}$$

---

[17] The NBER one-digit patent classes are: chemical, communications, medical, electric, mechanical, other, no class, and missing class. All results in this section are similar when using two-digit NBER patent classes, USPTO patent classes, or IPC classifications.

where C_{it} is the set of all patent classes in county i at time t. I construct this index using two-digit NBER patent classes, although results are similar with other patent class measures. Results are presented in Column 3 of Table A12. A new college causes concentration to increase, although the sign reverses after controlling for the number of patents granted in each county in Column 4 (since concentration is mechanically related to the number of patents, especially for small absolute numbers of patents), and neither estimate is statistically different from zero.

Table A12: Patent Classes

	Control for Patent Classes	By Patent Classes	Class Concentration (2-Dig.)	Class Concentration (2-Dig.)
College * PostCollege	0.262	0.205	4.012	-5.139
	(0.314)	(0.299)	(4.238)	(3.214)
PostCollege	-0.022	-0.005	-5.902*	-1.459
	(0.122)	(0.136)	(3.210)	(2.180)
Num. Pat.				0.730***
				(0.112)
Control for Distribution of Classes	Yes	No	No	No
Class FE	No	Yes	No	No
Class-Year FE	No	Yes	No	No
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1838-2010	1836-2010	1836-2010
Cnty-Year Obs.	16,984	77,562	20,900	20,900
# Counties	176	176	176	176
# Experiments	64	64	64	64
Mean of Dep. Var.	0.498	0.844	51.573	51.573
Adj. R-Sqr.	0.613	0.603	0.261	0.587

Notes: Column 1 includes a control for the fraction of patents in each NBER patent class. Column 2 estimates the results at the class-by-county-by-year level. Column 3 estimates the change in patent class concentration. Column 4 repeats the estimates in Column 3 but includes a control for the number of patents granted in each county. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D.I Patent Quality

As Trajtenberg (1990) makes clear, looking at raw patent counts without correcting for patent quality can produce misleading results. Ex ante, it is not clear whether patents in college counties should be expected to increase or decrease in average quality after establishing the college. On one hand, patents coming from more educated inventors might be expected to be of higher quality. On the other hand, more educated individuals, especially those trained in subjects like law, may have better access to the legal system and therefore patent more marginal inventions, leading to lower average quality. A third possibility is that the change in patenting is driven by shifts in the size of the population but not in the distribution of inventive abilities, in which case the distribution of patent qualities may not change at all. Following Hall et al. (2001) and Hall, Jaffe, and Trajtenberg (2005), I check whether the number of patent citations and citations per patent change in college counties relative to the runners-up after the establishment of a new college. I thank Enrico Berkes for providing lifetime citation counts for the universe of patents (see Berkes (2018)).

In Column 1 of Table A13, I show that the absolute number of patent citations in college counties increases by 82% relative to the runner-up counties after establishing a new college. This is a bit larger in magnitude than the percentage change in the total number of patents granted in college counties. The next three columns investigate changes in citations per patent after establishing a new college. Column 2 shows that citations per patent ($\frac{Citations_{it}}{Patents_{it}}$, where $Citations_{it}$ measures lifetime forward citations for all patents granted in county i in year t) declines dramatically in college counties relative to the runners-up after establishing a new college, although it is not statistically different from zero. In Column 2, I omit any counties with zero patents for which the number of patents in the denomination of $\frac{Citations_{it}}{Patents_{it}}$ is zero; in Column 3 I include these counties and code citations per patent to be zero in these cases, as well as including a dummy variable for zero patents. The coefficient is again negative but closer to zero in magnitude and not statistically significant. In Column 4, I also control for the distribution of patent classes in each county as in Column 1 of Table A12; this is due to the fact that some classes may inherently receive more citations than others. The coefficient is now very close to zero in magnitude and not statistically significant.

One possible reason for the lack of a large average effect, as noted above, is that more high quality patents in college counties could also be offset by more marginal patents. To simply check for this, I estimate whether the share of patents falling in the tails of the distribution of forward citations changes in the college counties relative to the runners-up following the establishment of a new college. Column 5 estimates the change in the fraction of a county's patents that fall below the 10th percentile of patents in terms of forward citations in each year. While I find a large (more than 70%) increase in the share of patents in the 10th percentile or below, the estimate is extremely noisy. In Column 6, I estimate the change in the fraction of patents falling in the 90th percentile of forward citations or above in each year; this coefficient is also large but extremely imprecisely estimated. Taking all the citation results together, there is no measurable change in citations per patent.

Unfortunately, patent citations are only consistently available beginning in 1947, making them a less-than-ideal measure when using historical patent data. I therefore use an alternative measure to gauge patent quality. As suggested in Kuhn and Thompson (2019), the length of a patent's first claim is an informative measure of a patent's scope, and hence its quality. A patent's claims formally define the legal scope of an invention. The first listed claim is the most broad. A very short first claim therefore indicates a patent that is very broad in scope, while a long claim indicates a patent that is narrow in scope. Kuhn and Thompson (2019) and Kuhn, Younge, and Marco (2017) argue that patent claim length is in fact more informative of patent quality than citation-based measures. Additionally, unlike patent citations, claims are recorded in the body of a patent for all patents granted in the U.S. from 1836 onward. I use the patent body text and claim counts from Enrico Berkes. I again thank Enrico Berkes for graciously providing this data.

In Column 1 of Table A14, I re-estimate the baseline regression specification using the average number of words in the first claim for all patents granted within each county in each year as the dependent variable. Column 2 uses the logged number of words in the first claim as the dependent variable. Neither measure is statistically significant and both are small in magnitude. Column 3 estimates the change in the share of patents at or below the tenth percentile of the first claim length distribution, representing the very broadest patents granted in a particular year. Column 4 estimates the change in the share of patents at or above the 90th percentile, the narrowest patents. Again, neither coefficient is large in magnitude. These results suggest that, while counties that receive a college gain more patents overall, there is no measurable change in patent quality.

D.J Results with Low-Quality Site Selection Experiments

In Table A15, I repeat the analysis in Columns 1-3 of Table 2 but include data from all colleges and counties for which runner-up sites can be identified. This includes the "low-quality" experiments as well as other runner-up counties in the high quality experiments that

Table A13: Patent Quality: Forward Citations

	log(Citations + 1)	Citations per Patent	Citations per Patent	Citations per Patent	Frac. Citations <10th Pcntl	Frac. Citations >90th Pcntl
College * PostCollege	0.818**	-0.657	-0.355	0.021	0.719	0.428
	(0.341)	(0.484)	(0.359)	(0.103)	(0.680)	(0.544)
PostCollege	-0.193*	0.268	-0.343	0.125	-0.353	-1.093
	(0.115)	(0.625)	(0.500)	(0.184)	(0.719)	(0.891)
Zero Pat. Dummy			-1.607***	0.449**		
			(0.222)	(0.211)		
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	33,660	16,984	33,660	33,660	13,763	13,763
# Counties	176	176	176	176	176	176
# Experiments	64	64	64	64	64	64
Mean of Dep. Var.	1.062	9.626	4.857	4.857	2.085	4.348
Adj. R-Sqr.	0.578	0.132	0.099	0.470	0.544	0.525

Notes: Column 1 estimates the change in the number of logged lifetime forward citations for all patents in a county. Column 2 estimates the change in the average citations per patent, omitting any counties with zero patents. Column 3 estimates the change in the average citations per patent, including a dummy variable for counties with zero patents. Column 4 re-estimates Column 3 but also controls for the distribution of patent classes. Column 5 estimates the change in the fraction of a county's patents that are at or below the 10th percentile of patents with respect to forward citations in each year. Column 6 estimates the change in the fraction of a county's patents that are at or above the 90th percentile of patents with respect to forward citations in each year. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table A14: Patent Quality: Claim Length

	Length of 1st Claim	log(Length 1st Claim)	Frac. 1st Claim <10th Pcntl	Frac. 1st Claim >90th Pcntl
College * PostCollege	-0.052	-0.026	0.108	-0.035
	(0.040)	(0.021)	(0.116)	(0.118)
PostCollege	0.091***	0.041**	-0.201*	0.183**
	(0.033)	(0.021)	(0.104)	(0.090)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	14,172	14,172	14,172	14,172
# Counties	176	176	176	176
# Experiments	64	64	64	64
Mean of Dep. Var.	46.432	3.744	0.078	0.095
Adj. R-Sqr.	0.432	0.503	0.017	0.013

Notes: Column 1 estimates the change in the average number of words in a patent's first claim in the college counties relative to the runner-up counties after the establishment of a college. Column 2 estimates the change in the average logged number of words in a patent's first claim. Column 3 estimates the change in the fraction of a county's patents that are at or below the 10th percentile of patents with respect to the length of first claim in each year. Column 4 estimates the change in the fraction of a county's patents that are at or above the 90th percentile of patents with respect to the length of first claim in each year. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

were nevertheless not as good as randomly assigned and so are excluded from the baseline sample. Instead of estimating Equation (1), I now estimate a triple-difference equation of the form

$$\begin{aligned} PatentMeasure_{it} = &\delta_1 College_i * HighQuality_i * PostCollege_{it} \\ &+ \delta_2 College_i * PostCollege_{it} + \delta_3 HighQuality_i * PostCollege_{it} \\ &+ \delta_4 PostCollege_{it} + County_i + Year_t + \epsilon_{it}. \end{aligned} \tag{5}$$

The indices mean the same as in the previous equations. Now $HighQuality_i$ is equal to one if county i is included in the original baseline sample (that consists of only the high quality experiments), and zero otherwise. In Column 1, I estimate Equation (5) using $\log(NumPat_{it} + 1)$ as the dependent variable. In Column 2, I use as the dependent variable the inverse hyperbolic sin of patenting, while in Column 3 the dependent variable is the alternative $\log(NumPat_{it})$ measure that includes a dummy equal to one if a county has zero patents in a particular year. In the new regression specifications, the coefficient of the triple-interaction term, $College_i * HighQuality_i * PostCollege_{it}$, measures how much larger the differences-in-differences estimator between high quality college and runner-up counties is compared to the differences-in-differences estimator between all college counties (high and low quality) and all runner-up counties (not just the high quality runners-up). This coefficient is negative and statistically significant at the 1% level in all columns, indicating that there is positive selection into the low quality college experiments; that is, the difference between the college and runner-up counties is smaller for high quality experiments than for counties not included in the baseline results. The coefficient on $College_i * PostCollege_{it}$

estimates the increase in patenting in *all* college counties relative to *all* runner-up counties after establishing a new college. The estimated coefficient is positive and significant, so the qualitative conclusions of the baseline specification in Table 2 are still true even if the low quality experiments are included, although the coefficients are a bit larger when attention is not restricted to the high quality experiments. The increase in patenting in high quality college counties over high quality runner-up counties after establishment of a new college (that is, the same quantity as estimated by the differences-in-differences term in Equation (1)) is given by adding the coefficient on the triple interaction term to the interaction term for all colleges in the post-college periods.[18] Combining these coefficients reveals that high quality college counties increase patenting by amounts slightly smaller than, but qualitatively similar to, the findings in Columns 1-3 of Table 2 (a 36% increase in Column 1, 44% increase in Column 2, and 27% increase in Column 3). All the combined coefficients are still statistically significant. The coefficient on $HighQuality_i * PostCollege_{it}$ estimates the change in patenting in high quality college and runner-up counties after the establishment of a college relative to low quality college and runner-up counties; these coefficients are positive and statistically significant, further suggesting that the high quality college and runner-up counties were more likely to be up-and-coming locations in terms of invention. Finally, the coefficient on $PostCollege_{it}$ has the same interpretation as before and simply measures the increase in patenting after establishment of a new college.

---

[18] Let $y = \log(NumPatents + 1)$. Then, the coefficient of interest is

$$\begin{aligned}(E[y_{College,HighQuality,PostCollege}]-E[y_{College,HighQuality,PreCollege}])& \\ -(E[y_{RunUp,HighQuality,PostCollege}] - E[y_{RunUp,HighQuality,PreCollege}])& \\ =[\delta_1+\delta_2+\delta_3+\delta_4]-[0]-[\delta_3+\delta_4]+[0]& \\ =\delta_1+\delta_2.&\end{aligned}$$

Table A15: Results with High and Low Quality College Site Selection Experiments

	log(Pat. +1)	arcsinh(Pat.)	Alt. log(Pat.)
College * HighQuality * PostCollege	-0.277***	-0.307***	-0.236***
	(0.098)	(0.109)	(0.080)
College * PostCollege	0.632***	0.744***	0.508***
	(0.138)	(0.169)	(0.101)
HighQuality * PostCollege	0.221***	0.276***	0.142***
	(0.065)	(0.080)	(0.042)
PostCollege	0.050***	0.064***	-0.031***
	(0.007)	(0.009)	(0.005)
Zero Pat. Dummy			-0.701***
			(0.006)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	2,426,699	2,426,699	2,426,699
# Counties	1,956	1,956	1,956
# Experiments	181	181	181
Mean of Dep. Var.	0.310	0.393	0.310
Adj. R-Sqr.	0.443	0.441	0.650

Notes: Column 1 estimates the level shift in patenting in college counties relative to all runner-up counties after establishment of a new college when the dependent variable is $log(Num.Patents + 1)$. The dependent variable in Column 2 is the inverse hyperbolic sin of patents. The dependent variable in Column 3 is $log(Num.Patents)$, with values replaced with 0 if $Num.Patents = 0$ and a dummy variable for zero patents included. Each coefficient is transformed into a percentage change in the dependent variable /100. When the model estimates changes in levels, the percentage change is calculated based on the baseline value of the independent variable in 1880. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

D.K Additional Results on Colleges and Population

I present several additional results relating to colleges and population. Because population variables are collected from the decennial U.S. population censuses, I first restrict the data to observations that occur only in the census years: 1840, 1850, 1860, etc. Thus the outcome variable is the log of the number of patents granted in the ten years closest to each census year. In Column 1 of Table A16, I reproduce the baseline result on patenting using only patenting in census years. The estimated coefficient is similar to the baseline coefficient estimated in Column 1 of Table 2, although not statistically significant, likely due to the smaller sample of years. Column 2 reproduces the results from Column 6 of Table 2 and shows the effect of a new college on logged county population.

In Column 3, I re-estimate Equation (1) but include $\log(TotalPop)$ as a control. Not surprisingly, county population is highly predictive of county patenting (a ten percent increase in population leads to a 6.2% increase in patenting). When including $\log(TotalPop)$, the coefficient on the interaction term of interest is only 39% of the baseline estimate, decreasing from 40% more patents per year in the baseline to 16% more patents per year. In Column 4, I remain agnostic about the functional form that population can take in the model, employing fractional polynomial regression as proposed by Royston and Altman (1994). I estimate a second degree polynomial, but results are similar with higher dimensions. I omit the coefficients on the polynomial terms for readability. When population is allowed to take a flexible form, the coefficient on the interaction term of interest drops even further, to only 25% of its baseline value. Thus, simply controlling for population in the baseline regression can explain about 60-75% of the observed increase in patenting. Moreover, in both cases the estimated effect of establishing a new college is not statistically significant at conventional levels, and so I cannot reject the null hypothesis that population can explain all of the observed increase in patenting in college counties after the establishment of a new college. Of course, these results should not be interpreted as causal; in the language of Angrist and Pischke (2009), population is a "bad control" for patenting.

If knowledge spillovers are larger when individuals can interact with alumni or college students, then simply controlling for population may be capturing the effect that migrants are endogenously sorting to places where these spillovers are largest. As a crude test of this, I check whether a marginal increase in population has a larger effect on patenting in college counties than in runner-up counties. Formally, I estimate

$$\begin{aligned}\log(NumPat_{it}+1) =& \delta_1 College_i * PostCollege_{it} + \delta_2 PostCollege_{it} + \delta_3 \log(TotalPop_{it}) \\ &+ \delta_4 College_i * PostCollege_{it} * \log(TotalPop_{it}) \\ &+ County_i + Experiment_i + Year_t + \epsilon_{it}. \qquad (6)\end{aligned}$$

Results are presented in Column 5 of Table A16. There is no evidence that increasing population increases patenting more in college counties (measured by δ_4). A 10% increase in population increases the differences-in-differences estimate by a statistically insignificant 1.5%. Results are similar when using other functional forms or semiparametric regressions for county population.

Appendix F.D shows heterogeneous treatment effects of establishing colleges on the basis of population at the time each college is established, among other dimensions of heterogeneity, and thus avoids challenges in interpreting ex post endogenous controls.

Table A16: The Effect of Colleges on Patenting when Controlling for Population

	log(Pat. + 1)	log(Total Pop.)	log(Pat. + 1)	log(Pat. + 1)	log(Pat. + 1)
College*PostCollege	0.402	0.486*	0.158	0.100	-0.769***
	(0.248)	(0.255)	(0.147)	(0.115)	(0.263)
PostCollege	-0.022	0.214*	-0.109*	-0.063	-0.980***
	(0.089)	(0.124)	(0.060)	(0.059)	(0.013)
log(Total Pop.)			0.620***		0.232***
			(0.099)		(0.057)
College * PostCollege * log(Total Pop.)					0.150
					(0.125)
Population Fract. Polynomials	No	No	No	Yes	No
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1840-2000	1840-2000	1840-2000	1840-2000	1840-2000
Cnty-Year Obs.	3,062	3,062	3,062	3,062	3,062
# Counties	176	176	176	176	176
# Experiments	64	64	64	64	64
Mean of Dep. Var.	0.694	9.759	0.694	0.694	0.694
Adj. R-Sqr.	0.626	0.721	0.688	0.734	0.711

Notes: Column 1 estimates the level shift in patenting in college counties relative to runner-up counties after establishment of a new college when the dependent variable is $log(Num.Patents + 1)$. Column 2 estimates the level shift in population in college counties relative to the runner-up counties after establishment of a new college when the dependent variable is $log(TotalPop.)$. The dependent variable for Columns 3-5 is $log(Patents + 1)$. Column 3 re-estimates Column 1 but includes a control for $log(TotalPop.)$. Column 4 re-estimates Column 1 but includes fractional polynomial controls for population. Column 5 estimates the effect of the level shift in patenting in college counties relative to runner-up counties after establishment of a new college when controlling for $log(TotalPop.)$ and interacting $log(TotalPop.)$ with a dummy for college counties, a dummy for post-college years, and the interaction term. Results are from census years only and $\log(Num.Patents + 1)$ measures the number of patents in the ten years closest to the census year. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

E Additional Results on Patenting by Alumni and Faculty

E.A Patenting by Alumni and Faculty Under Alternative Data Construction Assumptions

In this section, I reproduce results for the share of population and patents coming from alumni, faculty, and pre- and post-college others under several alternative data assumptions.

In the baseline results, shown in Table 3, I consider a patent to belong to an alumnus or faculty member if the name on a patent record matches to a name from the yearbook records, regardless of how many individuals in the census have the same name. This "John Smith problem" could substantially overstate the share of patents coming from alumni and faculty if a large share of patents belong to individuals with common names. In Table A17, I instead assign an alumnus or faculty $\frac{1}{N}$ of a patent if they share a name with N other individuals in the same county and census year. Perhaps surprisingly, this common name issue does not appear to substantially bias upwards the share of patents from alumni and faculty: under the new method, alumni and faculty account for about 4.9% of all patents instead of 5.3% in the baseline results.

In the baseline results, I consider to records to be a match if they have a bigram matching score of 0.8 or above.[19] In Table A18, I present matching results when requiring a ratio of 0.85 to consider two records to be a match. Shockingly, this slight increase in match strictness reduces the share of patents attributed to alumni and faculty to approximately zero. I am still able to match a similar number of patents to the census with this new criteria, so it appears that alumni and faculty names are less likely to exactly match census records than are patent records. A ratio of 0.8 is already quite liberal as a match cutoff; for instance, Sarada et al. (2019) use a cutoff of 0.9. The fact that I use a relatively liberal cutoff of 0.8 in the baseline results further suggests that the baseline results likely overstate the share of patents from alumni and faculty.

---

[19] The bigram score is calculated as the ratio of common two consecutive letter pairs in both the patent (or yearbook) record and census record to their average two consecutive letter pairs in both records. I compute this using Stata's reclink2 command (an extension of Blasnik (2007)); see Wasi and Flaaen (2015) for more details.

One of the concerns with using yearbook data, described in detail above in Section C, is that the yearbook data are not available for all years. This is especially problematic when computing the number of patents from alumni, since missing yearbooks will undercount the stock of possible alumni inventors. To adjust for this, I create an adjusted stock of potential alumni by interpolating student counts in missing yearbooks. In Table A19, I present results without making this adjustment for missing yearbook years. Not surprisingly, without this adjustment alumni and faculty account for only about 1.2% of the patents in their college counties. Table A20 presents results when using an alternative method to interpolate student and faculty counts for missing yearbook years. In Table A20, I use cubic splines to interpolate counts, whereas for the baseline results in Table 3 I use a linear interpolation. Results are similar in both tables.

Finally, I use an alternative method to determine whether a patentee was present in the college county at the time the college was established. In the baseline results, I consider a patentee to be present in the college county at the time the college was established if, as proposed by Ferrie (1996), an individual with a similar name appears in the census prior to the college establishment and the earlier record has the same race, gender, birthplace, and an appropriate age. In Table A21, I instead consider a patentee to be present in the college county at the time the college was established if a record with a similar name is present. Not surprisingly, this results in a much larger share of "pre-college others," since any patentee with a common name is likely to appear in multiple censuses.

Table A17: Patents by Alumni, Faculty, and Others: Multiple Matches

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	512.673	0.110	0.111	0.049	2.157
Faculty	19.013	0.004	0.000	0.000	0.000
Pre-College Others	1,159.196	0.248	0.411	0.183	3.549
Post-College Others	2,989.906	0.639	1.723	0.767	5.762

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county using fractional assignment of multiple patents. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

E.B Alumni and Faculty Patents in Counties with Only One College

If establishing a college spurs follow-up investment, including the creation of future colleges, then simply counting how many patents come from the alumni and faculty of an experiment college may be understating the direct effects of colleges. That is, alumni and faculty of other local colleges could be contributing a large number of local patents.

Table A18: Patents by Alumni, Faculty, and Others: Strict Matching Criteria

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	213.391	0.046	0.000	0.000	0.000
Faculty	4.890	0.001	0.000	0.000	0.000
Pre-College Others	1,277.891	0.277	0.688	0.228	5.382
Post-College Others	3,118.326	0.676	2.329	0.772	7.470

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county when using strict matching criteria. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A19: Patents by Alumni, Faculty, and Others: No Adjustment for Missing Yearbooks

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	47.029	0.007	0.036	0.012	7.649
Faculty	7.471	0.001	0.000	0.000	0.000
Pre-College Others	2,008.076	0.288	0.689	0.225	3.432
Post-College Others	4,910.721	0.704	2.340	0.763	4.764

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county without making any correction for missing yearbook years. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A20: Patents by Alumni, Faculty, and Others: Alternative Between-Yearbook Interpolation

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	673.511	0.097	0.119	0.039	1.768
Faculty	33.742	0.005	0.002	0.001	0.485
Pre-College Others	1,730.979	0.248	0.659	0.216	3.808
Post-College Others	4,535.065	0.650	2.277	0.745	5.021

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county when using a cubic spline to interpolate the number of students and faculty in missing yearbook years. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A21: Patents by Alumni, Faculty, and Others: Naive Matching Criteria

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	512.386	0.109	0.119	0.053	2.323
Faculty	19.013	0.004	0.001	0.000	0.474
Pre-College Others	1,159.337	0.248	0.411	0.183	3.541
Post-College Others	2,990.052	0.639	1.717	0.764	5.741

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county when using only first and last names to match individuals across censuses. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

To check for this, I repeat the results in Table 3 but exclude any counties that had an additional local college as of 1940. I compile a list of all currently operating non-experiment colleges in the yearbook college counties from the IPEDS data. Then, I manually look up the establishment date for each of these colleges and exclude any yearbook college counties that had another college established in 1940 or earlier. When excluding these colleges, the remaining yearbook colleges are: Auburn University, Iowa State University, Missouri University of Science and Technology, North Dakota State University, Texas Tech University, University of Arizona, University of Colorado, University of Nevada, University of New Hampshire, University of North Dakota, Utah State University, Virginia Tech University, and Washington State University.

Results are presented in Table A22. When restricting attention to these colleges, alumni and faculty together account for about 5.9% of all patents, only slightly larger than the 5.3% when using all of the yearbook colleges. Thus, it does not appear that missing patents from alumni and faculty of non-experiment local colleges are resulting in a substantial undercount of patents from alumni and faculty.

E.C The Role of Alumni and Faculty Today

In this section, I expand on the discussion in Section III.A, in which I argue that the conclusions about the role of alumni and faculty in local invention from the pre-1940 period still hold in the present. In particular, I explain the data construction and present results in detail. A challenge with the college yearbook data used in Section III is that it is only possible to match these to the decennial population censuses up to 1940. To check if the

Table A22: Patents by Alumni, Faculty, and Others: No Counties with Multiple Colleges

	Num. People	Share of Pop.	Num. Patents	Share Patents	Pat. per 10,000 Cap.
Alumni	115.217	0.084	0.023	0.059	2.031
Faculty	15.450	0.011	0.000	0.000	0.000
Pre-College Others	383.289	0.281	0.096	0.241	2.512
Post-College Others	851.012	0.623	0.279	0.700	3.283

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county, excluding all cases in which the college county had another local college established in 1940 or earlier. The first row lists statistics for alumni. The second row lists statistics for faculty. The third row lists statistics for other (non-alumni, non-faculty) individuals who were present in the college counties at the time the college was established. The fourth row lists statistics for other individuals who were not present in the college counties at the time the college was established. The first column lists the average number of people in each group per county. The second column lists the share of the county's total population belonging to each group. The third column lists the number of patents attributable to each group. The fourth column lists the share of the county's total patents attributable to each group. The fifth column lists the patenting rate for individuals in each group ($Num.Patents_j * 10,000/Num.Members_j$ for members of group j). Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

role of alumni and faculty in local invention is similar in recent years, one must construct proxies or find alternative ways of measuring the contribution of these groups.

E.C.1 Alumni Patenting

To construct measures of alumni patenting in post-1940 years, I use recent work by Bell, Chetty, Jaravel, Petkova, and Reenen (2019), who link patents to both IRS tax records and to alumni records used in Chetty, Friedman, Saez, Turner, and Yagan (2017). They provide data on the number of patents invented by college alumni for cohorts born from 1980-1984 and who attended college when they were 19-22 years old.[20] To make the Bell et al. (2019) consistent with the results from the college yearbooks, I restrict attention to only the colleges for which yearbook data is available.

---

[20] See `https://opportunityinsights.org/wp-content/uploads/2018/04/Inventors-Codebook-Table-3.pdf` for details on the construction of the alumni patenting data.

As when using the yearbook data, the Bell et al. (2019) provide the number of patents belonging to a subset of individuals who obtained their degrees in particular years. Instead, I am interested in the share of patents over a set of years for which all alumni are the inventors. To convert these data to the measure of interest, I divide the number of alumni patents granted in each year by the number of alumni that had graduated by that year to find the alumni patenting rate. I then use the IPEDS data to construct the stock of alumni in each college in each year.[21] Finally, I multiply the stock of alumni in each year by the alumni patenting rate to get the total number of alumni patents in each year. These steps are identical to those described in Section III and Appendix C.A.

One important difference between these alumni patenting counts and the pre-1940 counts of alumni patenting using the yearbook data in Section III is that these include patents by all alumni, *regardless of where they live.* In contrast, the results in Section III only show patenting by alumni in the counties of their alma maters. To adjust these results to reflect the degree of alumni geographic mobility, I use the results from Zolas, Goldschlag, Jarmin, Stephan, Owen-Smith, Rosen, Allen, Weinberg, and Lane (2015), who find that 87.3% of college graduates are living more than 50 miles from where they obtained their degree. I therefore scale the number of alumni by 0.127, treating this as a rough proxy for the share of alumni who live in a county different from that of their alma mater. Of course, naively applying the Zolas et al. (2015) “headline” mobility number has a number of drawbacks. First, it is an average over all colleges. Second, even within a college, the most talented and inventive alumni may also be the most mobile, or conversely, the most likely to remain in place to take advantage of their college’s resources. While the magnitude, or even the sign, of this bias is unclear, scaling the number of alumni by 0.127 is likely acceptable for the rough back-of-the-envelope nature of this calculation.[22]

---

[21] See `https://nces.ed.gov/ipeds/`.

With all these adjustments in mind, the alumni account for about 14.5% of all patents in the counties of their alma maters from 1996-2014. This is only slightly larger than the pre-1940 share in Table 3. Even under alternative assumptions, alumni still account for less than a quarter of all patents in college counties.

E.C.2 University-Assigned Patents

A natural substitute for faculty names from the yearbooks is to examine patents that are assigned to a particular college or university; in fact, this is the measure used in the sizable literature on university patenting (e.g., Mowery and Sampat (2001), Mowery and Ziedonis (2002), Sampat (2006)). Note that these two measures are not perfect substitutes. Linking patents to the names of faculty members included in the college yearbooks will capture all inventions by faculty members, even if they were not conducted using university resources or were conducted in a time period when a university did not require its staff to assign all inventions to the college. University-assigned patents, on the other hand, will capture all patents invented by individuals as part of their work for the university, even if these individuals were not tenure-line faculty and would be unlikely to be listed in a college yearbook.

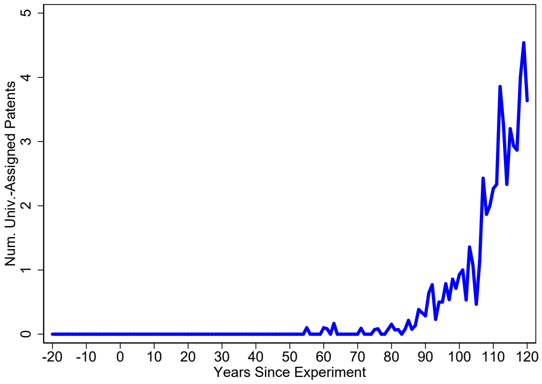
To create a list of university-assigned patents, for every patent granted in a college county, I check the assignee name for the words "College," "University," "Institute," or any of their common abbreviations. Note that this will capture *all* university-assigned patents in a county, and not just the patents assigned to colleges in my sample. So, for example, I include patents assigned to both Georgia Institute of Technology and Emory University in Fulton County, GA. This is done to minimize the risk of omitting a sample college's patents because of alternative ways in which the college name is written on a patent (for instance, the assignee may be the name of the university but may also include the name of a particular school or department within the university, may be assigned to the entire university system rather than a particular campus, etc.). But this decision likely overstates the number of patents belonging to faculty members of a particular college.

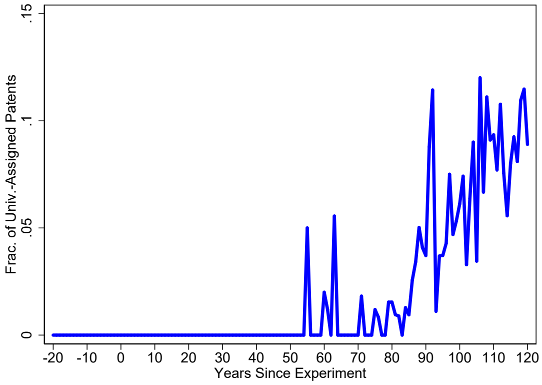
---

[22] I also calculate results when scaling the number of alumni patents by 0.22, which is the share of alumni living in the same state as their college according to Zolas et al. (2015). This should be thought of as an upper bound on the share of alumni remaining in the county of their alma mater. Under this alternative assumption, alumni still account for only 25% of all patents in the college county.

Results are presented in Figure A6. To make these results as comparable as possible to the results on faculty patenting in Section III, I present these results only for the college experiments for which yearbook data is available, although the results are similar for the sample of all colleges. The number of university-assigned patents, presented in Panel (a), has been increasing in recent decades. As a sanity check, I confirm that college counties received essentially zero university-assigned patents in the years before the college was established. Consistent with Sampat (2006), university-assigned patents began to rise in absolute terms in the decades before the passage of the Bayh-Dole Act in 1980 and have continued to increase in recent years, while university patenting was exceptionally rare before 1940; results by calendar year, rather than year since the college was established, are available upon request. While the number of university-assigned patents is growing rapidly in recent decades, the number of overall patents in college counties is growing nearly as quickly, so that the share of university-assigned patents grows only modestly; with the exception of one outlier year, university patents never account for more than 20% of all patents in college counties in any given year, and on average from 1996 to 2014 they account for only 4.5% of patents in college counties.

Figure A6: University-Assigned Patents





(a) Num. Patents

(b) Share of Patents

Notes: The x-axis shows the number of years since the college experiment. The year of the establishment of the new college is normalized to 0. Everything left of 0 shows pre-college results; everything to the right shows post-college results. In Panel (a), the y-axis shows the number of patents that list a college or university as an assignee in a college county. In Panel (b), the y-axis shows the share of total patents in the college counties that list a college or university as an assignee. Data are for college experiments for which yearbooks are available.

F Additional Results Investigating the Indirect Channels

F.A Most Common Inventor Occupations by Decade

Table A23 lists the top ten occupations for inventors, along with the share of inventors in each occupation, for the census years 1900-1940. Most common occupations are based on all inventors in the U.S. (not just inventors in college and runner-up counties) matched to the decennial population censuses in Sarada et al. (2019). These results reflect the democratization of invention (Khan, 2005) in the early years with the prevalence of skilled craftsmen (e.g., machinists, carpenters, painters) and the increasing specialization, professionalization, and technical skills needed to invent as exemplified by the growing role of engineers and managers.[23]

Table A23: Sensitivity to Different Types of Consolation Prizes

	1900		1910		1920		1930		1940	
	Occ.	%	Occ.	%	Occ.	%	Occ.	%	Occ.	%
1	CLERK	25.4	MANUFACTURER	11.7	CLERK	32.5	ENGINEER	26.8	SALESMAN	13.9
2	LABORER	9.59	LABORER	10.8	MANUFACTURER	8.52	MANAGER	8.64	MANAGER	12.2
3	MERCHANT	7.05	SALESMAN	6.80	LABORER	6.03	LABORER	6.88	OPERATOR	7.76
4	SHOEMAKER	4.75	OPERATOR	6.46	SALESMAN	5	SALESMAN	6.59	LABORER	7.15
5	MACHINIST	3.44	DRIVER	5.26	OPERATOR	4.90	CLERK	4.51	CLERK	5.28
6	DRIVER	3.42	CARPENTER	4.01	MACHINIST	3.49	OPERATOR	4.34	DRIVER	4.34
7	CARPENTER	2.88	CLERK	3.90	DRIVER	3.16	DRIVER	3.30	MECHANICAL ENGINEER	2.16
8	PAINTER	1.63	MACHINIST	3.64	CARPENTER	2.46	CARPENTER	2.14	MECHANIC	1.99
9	STUDENT	1.63	ENGINEER	2.52	ENGINEER	1.42	MACHINIST	1.86	PAINTER	1.86
10	ENGINEER	1.33	PAINTER	2.08	FOREMAN	1.33	PAINTER	1.78	CARPENTER	1.60

Notes: The ten most common occupation codes for patentees matched to the 1900, 1910, 1920, 1930, and 1940 decennial population censuses from Sarada et al. (2019), along with the percentage of inventors belonging to each occupation code in each year.

F.B Additional Consolation Prize Results

In this section, I present additional comparisons between college and consolation prize counties. In Column 1 of Table A24, I re-estimate the baseline consolation prize specification (Column 2 from Table 5) while controlling for logged population. In this case, the difference between the college and consolation prize counties is still statistically insignificant and close to zero in magnitude. Of course, population is an endogenous outcome variable, so this specification must be interpreted with caution.

Next, I show that the results are largely insensitive to the kinds of alternative institutions that are considered to be consolation prizes. In Column 2 I exclude any consolation prize counties that are educational institutions. This includes institutions like reform schools, which is a consolation prize for Weber County following the establishment of Utah State University in Cache County, UT. In Column 3, I again exclude educational institutions while also controlling for logged county population. The estimates are larger than the consolation prize results in Column 2 of Table 5 but still smaller than the baseline estimates. In Columns 4 and 5, I exclude any medical-related institutions, namely hospitals and insane asylums. In thee columns, the coefficient is not only statistically insignificant, but it is actually less than zero in magnitude. In all columns in Table A24, results are statistically indistinguishable from zero. One caveat is that restricting attention to particular types of consolation prize counties reduces the number of counties included in the regressions, reducing statistical power.

---

[23] The disappearance of "Engineers" (with an "occ1950" code of 16) in 1940 but the appearance of "Mechanical Engineer" ("occ1950" code of 460) suggests increasing specialization among a particularly important high-skilled occupation (Maloney and Caicedo, 2017). See `https://usa.ipums.org/usa/volii/occ_ind.shtml` for more information on the construction of occupation codes.

An additional concern is that, because the consolation prize counties were considered suitable sites to receive a college and because receiving a consolation prize institution induced local population growth, the consolation prize counties may be quite likely to receive a college of their own. If this were the case, then the presence of these colleges may be driving the observed similarity between the college and consolation prize counties. To assuage these concerns, I compile a list of all currently operating colleges in consolation prize counties from the IPEDS data. Then, I manually look up the establishment date for each of these colleges and exclude consolation prize counties from the analysis for any years after these consolation prize county colleges were established. I also omit the experiment college counties for any years in which every consolation prize county has a college.[24] I present results of this analysis in Table A25. Consistent with the analysis in Appendix D.B, consolation prize counties did indeed often get a college of their own at some point, although these did not typically appear very close in time to the opening of the focal college in the treated college county. Omitting consolation prize counties for any years after the consolation prize counties receive a college reduces the number of county-year observations by about 40%. The estimated coefficients are very similar even when excluding years in which consolation prize counties have colleges, and if anything the college counties perform even worse relative to the consolation prize counties in Table A25 than in Table 5.

---

[24] For example, Burleigh County, North Dakota, establishes Bismarck State College in 1939 and so is omitted from the analysis for all years 1939 and onward. In some cases, consolation prize counties establish colleges at the same time or even earlier than the college counties, as in the case of Stutsman County, North Dakota, which establishes the University of Jamestown in 1883, the same year that the University of North Dakota and North Dakota State were established. In these cases, the consolation prize county is omitted for all years.

Table A24: Sensitivity to Different Types of Consolation Prizes

	Control for Pop.	No Educ.-Related	No Educ.-Related	No Medical	No Medical
College * PostCollege	0.063	0.329	0.125	-0.031	-0.190
	(0.265)	(0.355)	(0.297)	(0.263)	(0.263)
PostCollege	-0.119	-0.077	-0.291*	0.299	0.061
	(0.216)	(0.176)	(0.160)	(0.255)	(0.373)
log(Total Pop.)	0.527**		0.517*		0.449
	(0.243)		(0.286)		(0.353)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	3,890	4,374	3,720	4,002	3,414
# Counties	27	25	25	23	23
# Experiments	13	13	13	13	13
Mean of Dep. Var.	0.400	0.252	0.252	0.274	0.274
Adj. R-Sqr.	0.645	0.509	0.534	0.528	0.540

Notes: Results omitting different types of consolation prizes. The dependent variable in all columns is $log(Patents + 1)$. In Column 1, I compare college to consolation prize counties while controlling for logged county population. In Column 2, college counties are compared to consolation prizes that exclude any type of educational institutions. In Column 4, college counties are compared to consolation prizes that exclude state capitals. In Column 6, college counties are compared consolation prizes that exclude medical institutions, namely asylums and hospitals. Columns 3, 5, and 7 repeat Columns 2, 4, and 6, respectively, but include a control for $log(Population)$. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table A25: Omit Consolation Prize Counties with Colleges

	State Univs.	Cons. Prize	No Cons. Prize
College * PostCollege	0.542**	-0.048	0.612**
	(0.226)	(0.403)	(0.249)
PostCollege	-0.023	-0.145	-0.008
	(0.102)	(0.217)	(0.110)
log(Total Pop.)	Yes	Yes	Yes
	Yes	Yes	Yes
County FE	1836-2010	1836-2010	1836-2010
Year FE			
Year Range			
Cnty-Year Obs.	17,283	2,030	15,253
# Counties	115	24	91
# Experiments	42	12	33
Mean of Dep. Var.	0.361	0.110	0.415
Adj. R-Sqr.	0.541	0.535	0.539

Notes: Results excluding consolation prize counties for years in which they have a college of their own. The dependent variable in all columns is $log(Patents + 1)$. Column 1 estimates the baseline level shift in patenting in college counties relative to runner-up counties after establishment of a new college for the subset of experiments that could potentially involve consolation prizes. Column 2 compares college counties to only runner-up counties that receive a consolation prize. Column 3 compares college counties to only runner-up counties that receive a consolation prize while controlling for $log(Population)$. Column 4 repeats Column 1 but excludes all counties that receives a consolation prize. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

F.C Heterogeneity by College Type

I further break down the results of different types of colleges on patenting. In Column 1 of Table A26, I control for logged county population, since different types of colleges may drive migration in different ways. After controlling for population, the coefficient is similar for both types of colleges, the coefficient for classical colleges is still larger in magnitude, and I cannot reject the null that neither type of college has any effect outside their effects on population. I stress that population is affected by the treatment, and so this specification should not be interpreted as causal. In Column 2 I show how patenting differs between practical and classical colleges, using an alternative classification of practical and classical than described in Section IV.C. Here, a college is considered a practical college if it is a land grant college, technical school, or military academy. Classical colleges are normal schools, other private and public colleges, and HBCUs. The difference between practical and classical colleges is qualitatively the same when the alternative definitions are used as in the baseline college type results presented in Table 5.

I also compare differences between each of the seven types of colleges: land grants, technical schools, military academies, normal schools, HBCUs, other public colleges, and other private colleges. Unfortunately, as Table 1 shows, there is only a small number of several types of experiments and so insufficient power to identify differences. Even simply comparing coefficients, however, paints a picture that does not conform to the naive intuition that colleges that focus on more practical skills should cause larger increases in patenting. For example, normal schools and land grant colleges produce nearly the same increase in patenting, while the former is focused on training primary and secondary school teachers and the latter has an explicit focus on very practical fields such as agriculture and machinery. These results are available upon request.

I next compare all public schools to all private schools. This involves reclassifying colleges, as some of the types described above may include both public and private colleges. For instance, the HBCUs may be either public or private. Cornell University, while officially New York's land grant university, is a private institution. I interact dummy variables for public or private status with the estimated college effect and display the results in Column 3. I find that public colleges have a large positive effect on patenting, while the effect for private colleges is larger in magnitude but not statistically different from zero. In Column 4, I control for logged county population, since public colleges may simply be larger and hence cause more population growth. Indeed, after controlling for population, the effect of both types of colleges are smaller and similar in magnitude, with neither individually statistically significant. As above, population is an endogenous outcome.

In Column 5, I check how the estimated treatment effect varies by college quality. Unfortunately, reliable data on college quality does not exist for most of each college's history. Instead, I proxy lifetime college quality with the 2018 national universities rankings in the U.S. New and World Reports (`https://www.usnews.com/best-colleges/rankings/national-universities`). This is problematic because current college rankings may be due in part to college's past patenting performance, but the measure may still be informative if rankings are highly persistent over time. I split colleges into four groups: those ranked 1-75, those ranked 76-150, those ranked 151-225, and those that do not have a 2018 U.S. News ranking. The estimated coefficient declines monotonically moving from the highest to lowest quality schools in the U.S. News rankings. These results suggest that higher quality colleges lead to more local patenting, although there are a relatively small number of experiments in each bin and no estimate is individually statistically significant at the 5% level. It may be the case that better colleges are larger, and it is the size of the institution that drives patenting rather than measures of quality. To try and account for this, in Column 5 I control for logged county population. The coefficients no longer decline monotonically, the coefficients are all smaller in magnitude and only the top ranked schools have a coefficient that is statistically significant at conventional levels. In sum, conclusions about the effect of college quality on local patenting are sensitive to the specification.

F.D Other Types of Heterogeneity

In this section, I examine the heterogeneity of the results along a number of additional, non-college related dimensions. In particular, I investigate whether the estimated effect of establishing a college on local invention systematically varies with preexisting county conditions.

For each year, I rank each county's position in the distribution of all U.S. counties for a number of characteristics, R_{it}. For the college counties, I then create a distribution of ranks in the last census year before each college was established, call it $R^c_{it^*_i}$, where the superscript c denotes college counties and t^*_i is the last census before college i is established. I separate the college experiments depending on whether the college county is above the 75th percentile or below the 25th percentile of the $R^c_{it^*_i}$ distribution. The purpose of this exercise is to ensure that counties aren't recorded as, for instance, being in the top quartile of the population distribution just because the college was established late, after several decades of population

Table A26: Additional Results by College Type

	College Types	Alt. Practical vs. Classical	Public vs. Private	Public vs. Private	College Rank	College Rank
Practical College Interaction	0.142	0.285				
	(0.178)	(0.205)				
Classical College Interaction	0.295	0.636				
	(0.316)	(0.441)				
log(Total Pop.)	0.596***			0.605***		0.601***
	(0.098)			(0.100)		(0.096)
All Public Colleges			0.477**	0.207		
			(0.220)	(0.151)		
All Private Colleges			0.507	0.188		
			(0.506)	(0.406)		
Rank 1-75					0.780*	0.594**
					(0.430)	(0.285)
Rank 76-150					0.427	0.157
					(0.279)	(0.254)
Rank 151-225					0.414	0.075
					(0.253)	(0.232)
Unranked					0.419	0.148
					(0.492)	(0.294)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	31,210	33,660	33,660	31,210	33,660	31,210
# Counties	176	176	176	176	176	176
# Experiments	64	64	64	64	64	64
Mean of Dep. Var.	0.498	0.498	0.498	0.498	0.498	0.498
Adj. R-Sqr.	0.566	0.522	0.521	0.566	0.524	0.567

Notes: Regression results by college type. The dependent variable is $log(Patents + 1)$. In Column 1, the effect of establishing a new college is estimated separately for practical and classical colleges while controlling for logged county population. In Column 2, the effect of establishing a new college is estimated separately for practical and classical colleges, using the alternate definition described in the text. In Column 3, the coefficient is the percentage increase in patenting caused by the college interacted with whether a college is public or private. Column 4 repeats the specification in Column 3 while also controlling for logged county population. In Column 5, the coefficient is the percentage increase in patenting caused by the college interacted with each college's rank according to the 2018 U.S. News and World Report rankings. Column 6 repeats the specification in Column 5 while controlling for logged county population. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

growth. Instead, I determine which counties had relatively higher or lower values of each characteristic relative to the other college counties at the time each college was established. Then, I estimate

$$
\begin{aligned}
PatentMeasure_{ijt} =& \delta_1 College_i * PostCollege_{it} * Above75thPct_i \\
& + \delta_2 College_i * PostCollege_{it} * Below25thPct_i \\
& + \delta_3 PostCollege_{jt} + County_i + Year_t + \epsilon_{ijt},
\end{aligned}
\tag{7}
$$

where $Above75thPct_j = 1$ if $R^c_{jt^*_j} \geq$75th-Percentile($R^c_{it^*_i}$) for some college j, $Below25thPct_j = 1$ if $R^c_{jt^*_j} <$25th-Percentile($R^c_{it^*_i}$), and the sample consists of only the cases when $Above75thPct_j = 1$ or $Below25thPct_j = 1$.

Table A27 presents results. Column 1 shows results for counties in the top and bottom quartiles of the population distribution. The coefficient is larger for counties in the top quartile at 50% more patents per year in the top quartile of college counties relative to their runner-ups, compared to 31% more patents per year in the bottom quartile of college counties relative to their runner-ups, although neither coefficients is statistically significant. Column 2 shows heterogeneity by the fraction of the county population living in an urban area. The coefficient is similar in both cases, with college counties in the top quartile having an imprecisely estimated increase of about 28% more patents per year relative to their runner-up counties, while the bottom quartile of college counties see an increase of 29% more patents per year relative to their runners-up. Column 3 shows that the bottom quartile of counties by prior patenting sees a substantially larger increase in patenting relative to their runners-up (42%, significant at the 5% level) than do the top quartile of counties (27%, statistically insignificant). Finally, Column 4 shows that the bottom quartile of counties as measured in the percentage of the county within 15 miles of access to a railroad line sees a substantially larger increase in patenting relative to their runners-up (140%, significant at the 1% level) than do the top quartile of counties (44%, statistically insignificant), although data on railroad access is not available for many counties.

Results examining heterogeneity by preexisting manufacturing and agricultural conditions are all imprecisely estimated. Results are also similar when dividing by the median instead of examining the top and bottom quartiles, although the inclusion of colleges just above and below the median cutoff makes the median results more difficult to interpret. These additional results are available upon request.

In sum, while these heterogeneity results do not present an unambiguously clear picture of when and where establishing a college will have a larger effect, they do suggest that the effect may be likely to be larger when an area is initially less developed, with a more rural, less inventive, and less connected population.[25] This conclusion is also consistent with the results in Section D.D, which finds that the effect of a college is larger when there is no preexisting college in the county, although the difference is modest. This is in turn consistent with the interpretation that the main channel through which colleges affect local invention is by acting as an initial anchor institution that drives future growth. Of course, these results cannot be interpreted as causal; it may be the case that more populous counties are able to fund larger colleges, for instance. Further exploring settings in which colleges have a larger or smaller effect is an important avenue for future work.

---

[25] The population results in Column 1 point in the opposite direction, although the difference in magnitude of the estimated effect between the top and bottom quartile is small.

Table A27: Heterogeneity by County Characteristics

	By Prev. County Pop.	By Prev. Frac. Urbanized	By Prev. County Pat.	By Prev. Access to Railroads
Above 75th Pct.	0.498	0.278	0.273	0.444
	(0.338)	(0.239)	(0.310)	(0.391)
Below 25th Pct.	0.309	0.293	0.422**	1.397***
	(0.363)	(0.204)	(0.167)	(0.048)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
Cnty-Year Obs.	16,070	21,819	28,935	1,050
# Counties	87	118	152	6
# Experiments	62	64	64	59
Mean of Dep. Var.	0.513	0.416	0.453	0.183
Adj. R-Sqr.	0.505	0.492	0.502	0.331

Notes: Regression results by preexisting county characteristics. The dependent variable is $log(Patents + 1)$. In Column 1, the effect of establishing a new college is estimated separately for counties in the top and bottom quartile of the distribution of college county populations, in Column 2 for the top and bottom quartiles by urbanization, in Column 3 by the top and bottom quartiles of patenting in pre-college years, and in Column 4 by the top and bottom quartile of access to railways. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

References

Acemoglu, D., U. Akcigit, and M. A. Celik (2014). Young, restless and creative: openness to disruption and creative innovation. NBER Working Paper.

Akcigit, U., J. Grigsby, and T. Nicholas (2017). The rise of American ingenuity: innovation and inventors of the golden age. Unpublished, University of Chicago.

Andrews, M. J. (2019a). Comparing historical patent datasets. Unpublished, NBER.

Andrews, M. J. (2019b). Historical appendix: How do institutions of higher education affect local invention? Evidence from the establishment of U.S. colleges. Unpublished, NBER.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Atack, J., M. Jaremski, and P. L. Rousseau (2014, December). American banking and the transportation revolution before the Civil War. *Journal of Economic History 74*(4), 943–986.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. V. Reenen (2019, May). Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics 134*(2), 647–713.

Berkes, E. (2018). Comprehensive universe of U.S. patents (CUSP): data and facts. Unpublished, Ohio State University.

Blasnik, M. (2007). RECLINK: Stata module to probabilistically match records. Statistical Software Components S456876, Boston College Department of Economics, revised 18 Jan 2010.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility report cards: the role of colleges in intergenerational mobility. Unpublished, NBER Working Paper No. 23618.

Ferrie, J. P. (1996, Fall). A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 29*(4), 141–156.

Goldin, C. and L. F. Katz (1999, Winter). The shaping of higher education: the formative years in the United States, 1890 to 1940. *Journal of Economic Perspectives 13*(1), 37–62.

Hall, B. H., A. Jaffe, and M. Trajtenberg (2005, Spring). Market value and patent citations. *RAND Journal of Economics 36*(1), 16–38.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: lessons, insights and methodological tools. NBER Working Paper.

Hausman, N. (2017). University innovation and local economic growth. Unpublished, Hebrew University.

Jones, B. F. (2009, January). The burden of knowledge and the "death of the renaissance man": is innovation getting harder? *Review of Economic Studies 76*(1), 283–317.

Jung, T. and O. Ejermo (2014, July). Demographic patterns and trends in patenting: gender, age, and education of inventors. *Technological Forecasting and Social Change 86*, 110–124.

Khan, B. Z. (2005). *The democratization of invention: patents and copyrights in American* *economic development, 1790-1920*. New York: Cambridge University Press.

Kinnear, D. L. (1972). *The first 100 years: a history of Virginia Polytechnic Institute and State University*. Blacksburg, VA: Virginia Polytechnic Institute Educational Foundation, Inc.

Kuhn, J. M. and N. C. Thompson (2019). How to measure and draw causal inferences with patent scope. *International Journal of the Economics of Business 26*(1), 5–38.

Kuhn, J. M., K. A. Younge, and A. C. Marco (2017). Patent citations reexamined: new data and methods. Unpublished, UNC.

Leggett, M. D. (1874). *Subject-matter index of patents for invention issued by the United States Patent Office from 1790 to 1873, inclusive*, Volume 1-3. Washington, DC: Government Printing Office.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014, July). Disambiguation and co-authorship networks of the U.S. Patent Inventor Database (1975-2010). *Research Policy 43*(6), 941–955.

Logan, T. D. and J. M. Parman (2017, March). The national rise in residential segregation. *Journal of Economic History 77*(1), 127–170.

Maloney, W. F. and F. V. Caicedo (2017). Engineering growth: innovative capacity and development in the Americas. Unpublished, IZA.

Marco, A. C., M. Carley, S. Jackson, and A. F. Myers (2015, June). The USPTO historical patent data files: two centuries of invention. Unpublished, USPTO Economic Working Paper No. 2015-1.

Miller, J. (2016a). Jim Shaw historical patent data. electronic correspondence, October 24, 2016.

Miller, J. (2016b). Jim Shaw historical patent data. electronic correspondence, November 1, 2016.

Mowery, D. C. and B. N. Sampat (2001, August). University patents and patent policy debates in the USA, 1925-1980. *Industrial and Corporate Change 10*(3), 781–814.

Mowery, D. C. and A. A. Ziedonis (2002, March). Academic patent quality and quantity before and after the Bayh-Dole Act in the United States. *Research Policy 31*(3), 399–418.

Petralia, S., P.-A. Balland, and D. L. Rigby (2016a, August). Data descriptor: unveiling the geography of historical patents in the united states from 1836-1975. *Nature: Scientific Data 3*, 1–14. Article number: 160074.

Petralia, S., P.-A. Balland, and D. L. Rigby (2016b). Historical patent dataset. `http://dx.doi.org/10.7910/DVN/BPC15W`, accessed September 15, 2016.

Royston, P. and D. G. Altman (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Journal of the Royal Statistical Society: Applied Statistics 43*(3), 429–467.

Sampat, B. N. (2006, July). Patenting and US academic research in the 20th century: the world before and after Bayh-Dole. *Research Policy 35*(6), 772–789.

Sarada, M. J. Andrews, and N. L. Ziebarth (2019, October). Changes in the demographics of American inventors, 1870-1940. *Explorations in Economic History 74*.

Trajtenberg, M. (1990, Spring). A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics 21*(1), 172–187.

US Patent and Trademark Office (2000). United States patent grants by state, county, and metropolitan area (utility patents, 1990-1999). Technology assessment and forecasting report, Office for Patent and Trademark Information, U.S. Patent and Trademark Office, Washington, DC.

US Patent and Trademark Office (2018). Patenting in U.S. metropolitan and micropolitan areas: breakout by organization: explanation of data. `https://www.uspto.gov/web/offices/ac/ido/oeip/taf/cbsa_asg/explain_cbsa_asg.htm`, accessed November 24, 2018.

Wallenstein, P. (1997). *Virginia Tech, land-grant university, 1872-1997: history of a school, a state, a nation.* Blacksburg, VA: Pocahontas Press, Inc.

Wasi, N. and A. Flaaen (2015). Record linkage using Stata: preprocessing, linking, and reviewing utilities. *The Stata Journal 15*(3), 672–697.

Zolas, N., N. Goldschlag, R. Jarmin, P. Stephan, J. Owen-Smith, R. F. Rosen, B. M. Allen, B. A. Weinberg, and J. I. Lane (2015, December). Wrapping it up in a person: examining employment and earnings outcomes for Ph.D. recipients. *Science 350*(6266), 1367–1371.