Online Appendix to:

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

Michael J. Andrews*

December 6, 2021

Contents

*University of Maryland Baltimore County. *Email*: mandrews@umbc.edu.

A Summary Statistics

Tables A1 and A2 list summary statistics for several patenting-related variables, as well as for all county-level variables included in the balance checks in Figure 2. For each variable, Table A1 lists summary statistics in 1880, which is the last decennial census year before the mean and median college in the sample is established and so gives a snapshot of the country around the time most of the site selection experiments take place. Table A2 lists summary statistics over the entire sample period. Both views are informative given the dramatic secular increase in patenting, population, and urbanization that took place between 1836 and 2010. Some of the variables, namely the area of a county within 15 miles of a railroad and the fraction attending school, are not available in the 1880 census and are only available from a few years.

Table A1: Summary Statistics, 1880

	Mean	S.D.	1880 Min.	Median	Max.
Number of Patents	2.317	31.678	0.000	0.000	1,336.000
Any Patents	0.283	0.450	0.000	0.000	1.000
Population	20,054.201	47,086.173	0.000	12,687.000	1,296,873.000
Patents per Capita	0.427	3.544	0.000	0.000	275.229
Fraction Urban	0.083	0.175	0.000	0.000	1.000
Fraction Interstate Migrant	0.485	0.286	0.004	0.512	1.000
Manufacturing Output	2,306,597.212	16,719,234.443	0.000	173,526.000	482,030,784.000
Agricultural Output	930,333.955	991,924.950	0.000	612,952.000	9,320,202.000

Notes: Mean, standard deviation, minimum, median, and maximum values for several patent-related and other county-level economic and demographic variables. All data is from 1880.

Table A2: Summary Statistics, All Years

	All Years				
	Mean	S.D.	Min.	Median	Max.
Number of Patents	10.291	97.562	0.000	0.000	10,506.000
Any Patents	0.406	0.491	0.000	0.000	1.000
Population	50,621.945	224,521.828	0.000	16,532.000	9,758,256.000
Patents per Capita	0.740	7.352	0.000	0.000	2,903.226
Fraction Urban	0.150	0.271	0.000	0.000	13.564
Fraction Interstate Migrant	0.414	0.283	0.002	0.364	1.000
Manufacturing Output	11,000,600.698	81,922,990.756	0.000	408,441.000	2838989568.000
Agricultural Output	2,460,670.855	5,313,644.425	0.000	1,077,056.000	156,962,336.000
Area within 15 miles of RR	521.304	327.432	0.059	474.379	2,623.655
Fraction Attending School	0.184	0.082	0.000	0.208	0.548

Notes: Mean, standard deviation, minimum, median, and maximum values for several patent-related and other county-level economic and demographic variables. Data covers the entire sample period from 1836 to 2010.

B More Information on the College Site Selection Experiments

Table A3 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.[1] 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 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 (2021*a*).

Table A4 list the number of patents associated with each college site selection experiment.

[1]Table A3 is also included as an appendix to the main paper.

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 38 of the 63 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.

B.A Comparing Sample to Non-Sample Colleges

To compare the sample colleges to the 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 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

Table A3: List of College Site Selection Experiments

	College	County	State	Runner-Up Counties	Experiment Year	College Type
1	University of Missouri	Boone	Missouri	Howard; Cooper; Saline; Cole; Callaway	1839	Other Public
2	University of Mississippi	Lafayette	Mississippi	Harrison; Rankin; Montgomery; Attala; Winston; Monroe	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	Middlesex; Burlington; Essex	1855	Normal School
6	University of California Berkeley	Alameda	California	Napa; Contra Costa	1857	Land Grant
7	Iowa State University	Story	Iowa	Polk; Marshall; Jefferson; Tama; Hardin	1859	Land Grant
8	University of South Dakota	Clay	South Dakota	Yankton; Bon Homme	1862	Other Public
9	University of Kansas	Douglas	Kansas	Shawnee	1863	Other Public
10	Lincoln College (IL)	Logan	Illinois	Macon; Edgar; Warrick	1864	Other Private
11	Cornell University	Tompkins	New York	Seneca; Onondaga; Schuyler	1865	Land Grant
12	University of Maine	Penobscot	Maine	Sagadahoc	1866	Land Grant
13	University of Wisconsin	Dane	Wisconsin	Fond du Lac	1866	Land Grant
14	University of Illinois	Champaign	Illinois	Morgan; McLean	1867	Land Grant
15	West Virginia University	Monongalia	West Virginia	Greenbrier; Kanawha	1867	Land Grant
16	Oregon State University	Benton	Oregon	Marion	1868	Land Grant
17	Purdue University	Tippecanoe	Indiana	Marion; Hancock	1869	Land Grant
18	Southern Illinois University	Jackson	Illinois	Clinton; Jefferson; Washington; Perry; Marion	1869	Normal School
19	University of Tennessee	Knox	Tennessee	Rutherford	1869	Land Grant
20	Louisiana State University	East Baton Rouge	Louisiana	Bienville; East Feliciana	1870	Land Grant
21	Missouri University of Science and Technology	Phelps	Missouri	Iron	1870	Technical School
22	Texas A and M University	Brazos	Texas	Austin; Grimes	1871	Land Grant
23	University of Arkansas	Washington	Arkansas	Independence	1871	Land Grant
24	Auburn University	Lee	Alabama	Tuscaloosa; Lauderdale	1872	Land Grant
25	University of Oregon	Lane	Oregon	Polk; Linn; Washington	1872	Other Public
26	Virginia Polytechnic Institute	Montgomery	Virginia	Albemarle; Rockbridge	1872	Land Grant
27	University of Colorado	Boulder	Colorado	Fremont	1874	Other Public
28	University of Texas Austin	Travis	Texas	Smith	1881	Other Public
29	University of Texas Medical Branch	Galveston	Texas	Harris	1881	Technical School
30	North Dakota State University	Cass	North Dakota	Stutsman	1883	Land Grant
31	University of North Dakota	Grand Forks	North Dakota	Burleigh	1883	Other Public
32	University of Arizona	Pima	Arizona	Pinal	1885	Land Grant
33	University of Nevada	Washoe	Nevada	Carson City	1885	Land Grant
34	Georgia Institute of Technology	Fulton	Georgia	Greene; Clarke; Baldwin; Bibb	1886	Technical School
35	Kentucky State University	Franklin	Kentucky	Daviess; Boyle; Christian; Fayette; Warren	1886	HBCU
36	North Carolina State University	Wake	North Carolina	Mecklenburg; Lenoir	1886	Land Grant
37	University of Wyoming	Albany	Wyoming	Uinta; Laramie	1886	Land Grant
38	Utah State University	Cache	Utah	Weber	1888	Land Grant
39	Clemson University	Pickens	South Carolina	Richland	1889	Land Grant
40	New Mexico State University	Dona Ana	New Mexico	San Miguel	1889	Land Grant
41	University of Idaho	Latah	Idaho	Bonneville	1889	Land Grant
42	Alabama Agricultural and Mechanical University	Madison	Alabama	Montgomery	1891	HBCU
43	University of New Hampshire	Strafford	New Hampshire	Belknap	1891	Land Grant
44	Washington State University	Whitman	Washington	Yakima	1891	Land Grant
45	North Carolina A and T University	Guilford	North Carolina	Alamance; New Hanover; Durham; Forsyth	1892	HBCU
46	Northern Illinois University	DeKalb	Illinois	Winnebago	1895	Normal School
47	Western Illinois University	McDonough	Illinois	Schuyler; Warren; Hancock; Mercer; Adams	1899	Normal School
48	University of Nebraska at Kearney	Buffalo	Nebraska	Valley; Custer	1903	Normal School
49	Western Michigan University	Kalamazoo	Michigan	Allegan; Barry	1903	Normal School
50	University of Florida	Alachua	Florida	Columbia	1905	Land Grant
51	Georgia Southern College	Bulloch	Georgia	Tattnall; Emanuel	1906	Other Public
52	University of California Davis	Yolo	California	Solano	1906	Land Grant
53	East Carolina University	Pitt	North Carolina	Edgecombe; Beaufort	1907	Technical School
54	Western State Colorado University	Gunnison	Colorado	Mesa; Garfield	1909	Normal School
55	Arkansas Tech University	Pope	Arkansas	Sebastian; Franklin; Conway	1910	Technical School
56	Bowling Green State University	Wood	Ohio	Henry; Sandusky; Van Wert	1910	Normal School
57	Kent State University	Portage	Ohio	Medina; Trumbull	1910	Normal School
58	Southern Arkansas University	Columbia	Arkansas	Ouachita; Polk; Hempstead	1910	Other Public
59	Southern Mississippi University	Forrest	Mississippi	Hinds; Jones	1910	Normal School
60	Southern Methodist University	Dallas	Texas	Tarrant	1911	Other Private
61	Texas Tech	Lubbock	Texas	Nolan; Scurry	1923	Technical School
62	US Merchant Marine Academy	Nassau	New York	Bristol	1941	Military Academy
63	US Air Force Academy	El Paso	Colorado	Madison; Walworth	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 A4: 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	1618	960	659	0	0
2	University of Mississippi	1066	150	916	0	0
3	Eastern Michigan University	14202	11290	2912	3	1
4	Pennsylvania State University	3506	2120	1386	2	4
5	The College of New Jersey	67208	16877	50331	5	144
6	University of California Berkeley	62095	42813	19282	0	0
7	Iowa State University	7739	2180	5559	0	1
8	University of South Dakota	156	53	103	0	0
9	University of Kansas	2293	1045	1247	1	0
10	Lincoln College (IL)	3670	358	3312	9	17
11	Cornell University	16533	3194	13339	90	265
12	University of Maine	977	772	205	91	16
13	University of Wisconsin	10436	8803	1633	36	69
14	University of Illinois	4003	2522	1482	9	78
15	West Virginia University	2664	495	2169	11	0
16	Oregon State University	4762	3363	1400	0	2
17	Purdue University	21693	2597	19096	14	232
18	Southern Illinois University	1219	360	860	0	19
19	University of Tennessee	4659	4182	477	14	0
20	Louisiana State University	6792	6677	115	14	0
21	Missouri University of Science and Technology	538	419	119	0	2
22	Texas A and M University	1263	1068	195	0	0
23	University of Arkansas	720	654	67	0	2
24	Auburn University	1623	600	1023	0	6
25	University of Oregon	14039	2703	11336	0	4
26	Virginia Polytechnic Institute	1547	1145	402	0	11
27	University of Colorado	13319	13047	272	0	0
28	University of Texas Austin	27294	26512	782	37	12
29	University of Texas Medical Branch	48725	2582	46143	50	24
30	North Dakota State University	1194	998	196	0	0
31	University of North Dakota	643	275	368	0	0
32	University of Arizona	8725	8181	544	4	6
33	University of Nevada	4314	3767	547	7	3
34	Georgia Institute of Technology	13699	11838	1861	80	99
35	Kentucky State University	5237	115	5122	11	124
36	North Carolina State University	20644	14422	6222	26	30
37	University of Wyoming	797	267	530	6	8
38	Utah State University	2781	1279	1502	1	3
39	Clemson University	2535	915	1620	0	25
40	New Mexico State University	475	407	69	0	6
41	University of Idaho	1473	425	1048	0	0
42	Alabama Agricultural and Mechanical University	5075	4457	619	17	34
43	University of New Hampshire	1855	1273	582	56	85
44	Washington State University	1727	695	1032	3	2
45	North Carolina A and T University	13054	4801	8253	26	66
46	Northern Illinois University	9970	1434	8536	249	508
47	Western Illinois University	3210	402	2808	141	696
48	University of Nebraska at Kearney	378	248	130	26	22
49	Western Michigan University	9571	7869	1702	397	149
50	University of Florida	2892	2698	194	18	10
51	Georgia Southern College	140	89	51	0	14
52	University of California Davis	3716	2073	1643	42	55
53	East Carolina University	722	460	262	8	19
54	Western State Colorado University	1245	132	1113	21	49
55	Arkansas Tech University	695	133	563	9	64
56	Bowling Green State University	3181	1814	1367	75	417
57	Kent State University	7290	2645	4646	175	291
58	Southern Arkansas University	383	142	241	22	35
59	Southern Mississippi University	1020	300	720	0	55
60	Southern Methodist University	39872	29916	9956	198	182
61	Texas Tech	1355	1232	123	3	7
62	US Merchant Marine Academy	29831	24291	5540	659	2374
63	US Air Force Academy	10394	5449	4945	356	926

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.

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 Tables A5-A8, I show summary statistics for the high quality colleges, the low quality colleges, and the non-experimental colleges for the select years: 1870, 1890, 1900, and 1910; results for other years are available upon request. A non-experimental college is a college that is in neither a high quality nor low quality experiment in my sample.[2] I also show summary statistics for the subset of non-experimental colleges with a Carnegie Classification of R1 or R2, that is, to all institutions rated as having "very high" or "high" research activity, respectively. Data on the Carnegie Classification for each college is obtained from the IPEDS data. In each table, I show the mean, standard deviation, minimum, median, and maximum values for the number of students and number of faculty, and, when available, the number of graduate students, number of library volumes, and tuition. For most years, the high quality colleges are similar to the low quality colleges; in some years they have more students and faculty, and in some years they have fewer. For most years, both are larger on average than the non-experimental colleges and smaller than the subset of non-experimental colleges that currently have a Carnegie R1 or R2 classification. The changes over years reflects both changes in the composition of the sample as new colleges are established as well as the evolution of existing colleges.

To more fully show the entire distribution of colleges, in each panel of Figure A1, I plot the distribution of college characteristics of interests for the non-experimental colleges with green bars. All variables are residualized after controlling for year effects. For each variable,

[2]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.

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, although the high quality experiments are over-represented in the 10th decile of faculty while the low quality experiments are under-represented. 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.

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. Across all panels, the distribution of the sample colleges is fairly 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, although perhaps a bit smaller than, 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 universities but likely to overstate the effects of a college relative to the "typical" college established in the U.S.

B.B Additional Balance Checks

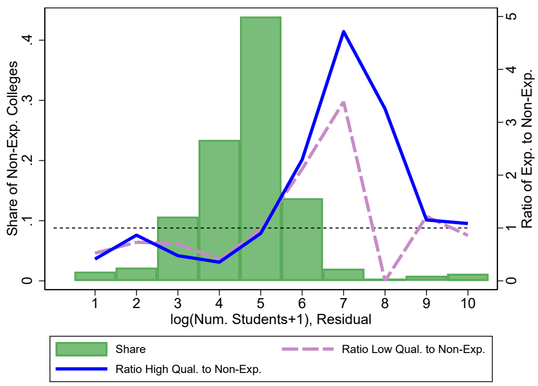
In Figure 2 in Section I.F, I compare college counties to runner-up counties along a number of observable dimensions and find that no individual dimension predicts treatment status. Here,

Table A5: Summary Statistics for Different Types of Colleges: 1870

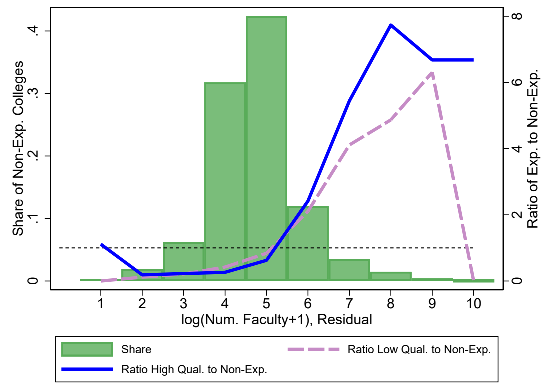
	N	Mean	S.D.	Min	Median	Max
			Num. Students			
High Quality Colleges	14	188.29	153.81	0	166	585
Low Quality Colleges	30	245.37	196.31	0	164	767
Non-Experiment Colleges	291	151.18	145.31	0	124	1,074
Non-Experiment Carnegie Colleges	4	192.50	152.20	76	140	414
			Num. Faculty			
High Quality Colleges	14	13.57	9.10	0	13	38
Low Quality Colleges	30	15.07	13.23	0	10	68
Non-Experiment Colleges	291	8.72	8.03	0	7	75
Non-Experiment Carnegie Colleges	4	17.25	8.46	10	17	26
			Library Volumes			
High Quality Colleges	14	10,285.71	13,487.48	0	5,000	38,000
Low Quality Colleges	29	11,162.66	11,568.75	0	6,000	37,967
Non-Experiment Colleges	290	3,376.16	5,114.32	0	1,253	34,000
Non-Experiment Carnegie Colleges	4	5,286.25	3,601.77	0	6,750	7,645
			Tuition			
High Quality Colleges	14	42.86	31.89	0	48	90
Low Quality Colleges	30	37.14	33.23	0	30	150
Non-Experiment Colleges	291	59.74	154.98	0	34	2,540
Non-Experiment Carnegie Colleges	4	29.75	23.81	4	28	60

Notes: Summary statistics for the high quality colleges, low quality colleges, non-experimental colleges, and non-experiments colleges that are classified as Carnegie R1 or R2 institutions, with each type of college on a different row. Columns show the mean, standard deviation, minimum, median, and maximum for each college type and college characteristic. The first set of four rows shows results for the number of students, the second set for the number of faculty, the third set for the number of graduate students, the fourth set for the number of library volumes, and the fifth set for tuition. Data are from the Commissioner of Education report in 1870.

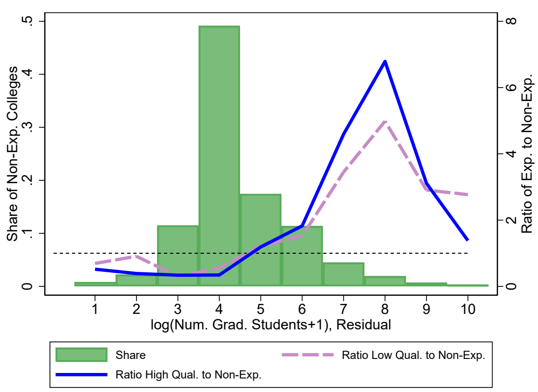
Figure A1: Compare Colleges In Sample to All Out of the Sample Colleges



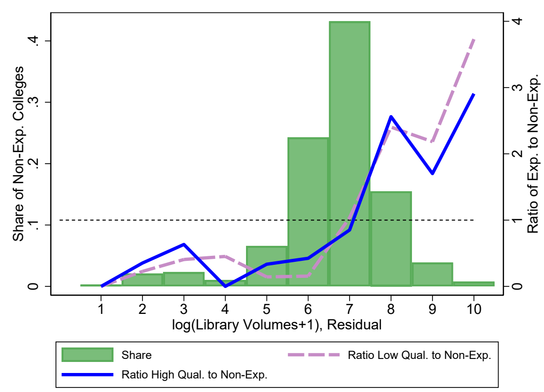
(a) log(Undergrad. Students)



(b) log(Faculty)



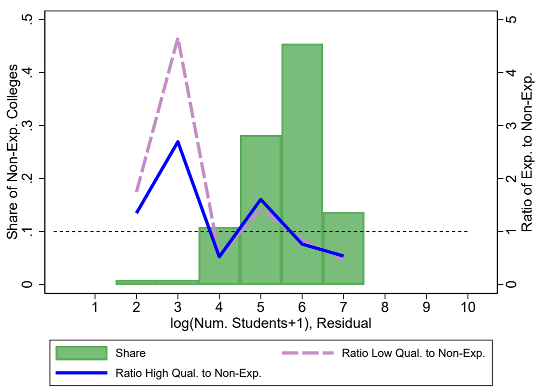
(c) log(Grad. Students)



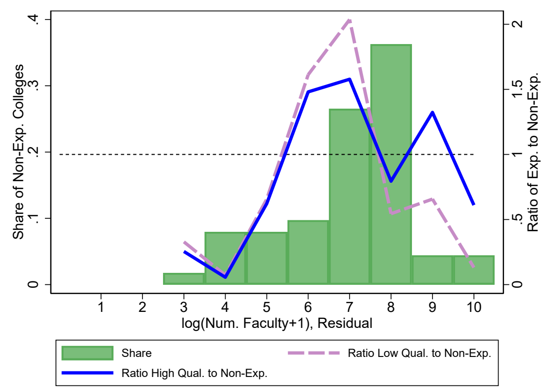
(d) log(Library Vols.)

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.

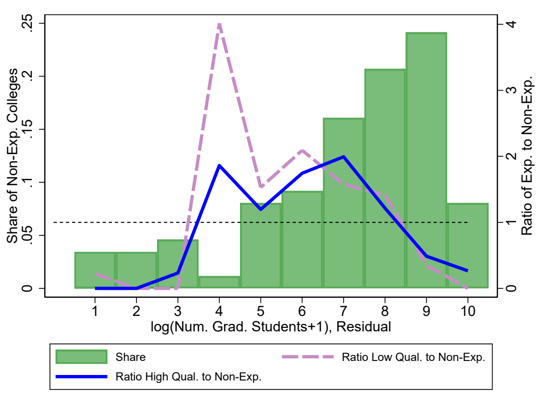
Figure A2: Compare Colleges In Sample to Out of the Sample Carnegie Research Institutions



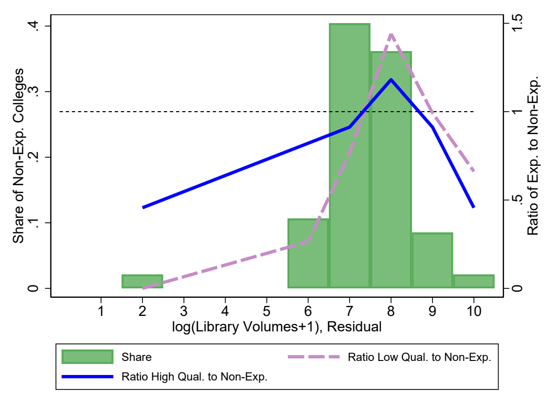
(a) log(Undergrad. Students)



(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.

Table A6: Summary Statistics for Different Types of Colleges: 1890

	N	Mean	S.D.	Min	Median	Max
			Num. Students			
High Quality Colleges	21	372.76	315.72	0	281	1,391
Low Quality Colleges	35	406.37	358.85	0	297	1,645
Non-Experiment Colleges	265	264.02	274.68	0	188	2,271
Non-Experiment Carnegie Colleges	10	789.50	694.68	70	548	2,271
			Num. Faculty			
High Quality Colleges	21	31.95	26.68	10	21	110
Low Quality Colleges	35	36.31	36.86	4	22	147
Non-Experiment Colleges	265	18.18	25.08	3	13	243
Non-Experiment Carnegie Colleges	10	84.60	71.08	15	66	243

Notes: Summary statistics for the high quality colleges, low quality colleges, non-experimental colleges, and non-experiments colleges that are classified as Carnegie R1 or R2 institutions, with each type of college on a different row. Columns show the mean, standard deviation, minimum, median, and maximum for each college type and college characteristic. The first set of four rows shows results for the number of students, the second set for the number of faculty, the third set for the number of graduate students, the fourth set for the number of library volumes, and the fifth set for tuition. Data are from the Commissioner of Education report in 1890.

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 a 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 A9. 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: logged patenting, logged

Table A7: Summary Statistics for Different Types of Colleges: 1900

	N	Mean	S.D.	Min	Median	Max
			Num. Students			
High Quality Colleges	27	803.48	862.22	84	387	3,337
Low Quality Colleges	42	769.48	850.93	92	468	3,413
Non-Experiment Colleges	298	351.77	538.01	21	214	5,225
Non-Experiment Carnegie Colleges	14	1,360.93	1,284.98	33	820	4,288
			Num. Faculty			
High Quality Colleges	27	67.11	74.15	12	43	327
Low Quality Colleges	42	67.21	83.62	8	34	359
Non-Experiment Colleges	298	26.00	41.89	3	15	483
Non-Experiment Carnegie Colleges	14	138.07	128.17	11	108	483
			Num. Grad. Students			
High Quality Colleges	27	36.19	54.02	0	10	205
Low Quality Colleges	42	40.71	85.93	0	6	433
Non-Experiment Colleges	298	10.27	64.78	0	0	1,003
Non-Experiment Carnegie Colleges	14	148.07	266.37	0	31	1,003
			Library Volumes			
High Quality Colleges	21	17,125.52	10,640.58	4,210	17,526	36,000
Low Quality Colleges	31	17,479.45	11,417.34	1,000	16,000	39,000
Non-Experiment Colleges	279	8,590.63	9,096.98	0	5,200	40,000
Non-Experiment Carnegie Colleges	6	20,967.00	11,684.68	4,000	20,300	37,202
			Tuition			
High Quality Colleges	27	23.60	33.36	0	11	121
Low Quality Colleges	40	46.10	41.79	0	36	192
Non-Experiment Colleges	257	37.55	42.03	0	25	306
Non-Experiment Carnegie Colleges	14	69.43	49.19	0	70	149

Notes: Summary statistics for the high quality colleges, low quality colleges, non-experimental colleges, and non-experiments colleges that are classified as Carnegie R1 or R2 institutions, with each type of college on a different row. Columns show the mean, standard deviation, minimum, median, and maximum for each college type and college characteristic. The first set of four rows shows results for the number of students, the second set for the number of faculty, the third set for the number of graduate students, the fourth set for the number of library volumes, and the fifth set for tuition. Data are from the Commissioner of Education report in 1900.

Table A8: Summary Statistics for Different Types of Colleges: 1910

	N	Mean	S.D.	Min	Median	Max
			Num. Students			
High Quality Colleges	37	994.68	1,093.89	60	544	4,896
Low Quality Colleges	52	1,100.83	1,325.95	124	592	5,422
Non-Experiment Colleges	380	417.45	631.61	21	252	7,028
Non-Experiment Carnegie Colleges	18	1,883.44	1,876.67	116	1,007	7,028
			Num. Faculty			
High Quality Colleges	37	112.57	138.95	8	61	652
Low Quality Colleges	52	100.65	120.79	8	55	486
Non-Experiment Colleges	380	37.41	61.99	3	20	618
Non-Experiment Carnegie Colleges	18	192.72	170.08	6	166	618
			Num. Grad. Students			
High Quality Colleges	37	39.65	80.31	0	9	372
Low Quality Colleges	52	66.46	201.87	0	5	1,367
Non-Experiment Colleges	380	14.23	96.12	0	0	1,638
Non-Experiment Carnegie Colleges	18	209.50	397.85	0	62	1,638
			Library Volumes			
High Quality Colleges	23	19,387.83	9,405.94	200	19,470	35,000
Low Quality Colleges	26	16,820.69	8,628.30	4,000	15,153	34,448
Non-Experiment Colleges	337	10,314.07	9,592.00	200	7,000	40,000
Non-Experiment Carnegie Colleges	7	17,146.86	12,184.68	7,000	10,000	36,000
			Tuition			
High Quality Colleges	35	29.60	21.02	7	24	106
Low Quality Colleges	50	59.91	43.55	6	52	171
Non-Experiment Colleges	342	65.22	57.76	4	50	588
Non-Experiment Carnegie Colleges	18	94.18	54.37	13	85	184

Notes: Summary statistics for the high quality colleges, low quality colleges, non-experimental colleges, and non-experiments colleges that are classified as Carnegie R1 or R2 institutions, with each type of college on a different row. Columns show the mean, standard deviation, minimum, median, and maximum for each college type and college characteristic. The first set of four rows shows results for the number of students, the second set for the number of faculty, the third set for the number of graduate students, the fourth set for the number of library volumes, and the fifth set for tuition. Data are from the Commissioner of Education report in 1910.

population, and the fraction urbanized.[3] The F-test statistic for the joint significance of these included regressors is 0.847, which is statistically insignificant. 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 fewer observations; the coefficients are again not jointly significant.[4] 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 they evolve similarly as well. In Panel (a), I plot logged county population for several decades

[3] Even here, population and urbanization data is missing for four experiments.

[4] 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.

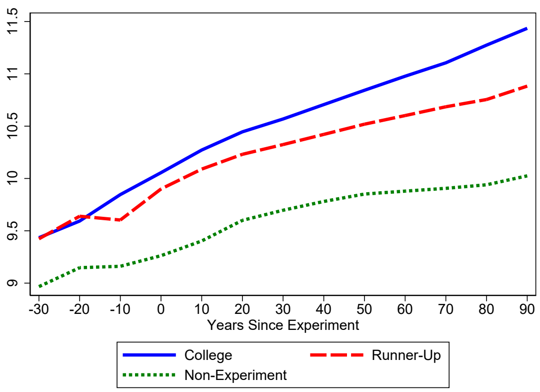
both before and after the establishment of a the new college in the college, runner-up, and non-experimental counties. Panel (b) plots the fraction of the county population that lives in an urban area. Panel (c) plots the logged farm output. Finally, Panel (d) plots logged manufacturing output. Plots 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 before the college is established.

Table A9: Tests for Joint Significance of Covariates Predicting Whether a County Receives a College

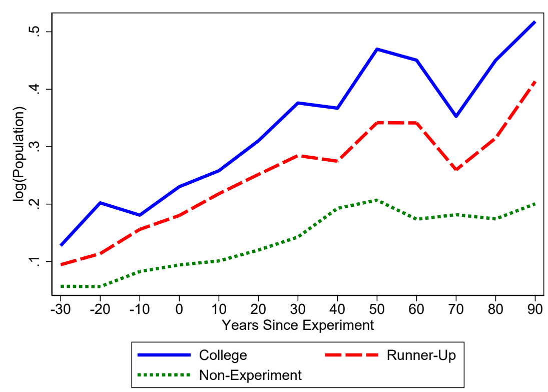
	Linear Probability	Logit	Linear Probability	Logit
log(Patents + 1)	-0.011	-0.060	0.013	0.032
	(0.052)	(0.234)	(0.090)	(0.396)
log(Population)	0.036	0.195	0.084	0.394
	(0.039)	(0.204)	(0.128)	(0.565)
Fraction Living in Urban Areas	0.107	0.448	-0.265	-1.443
	(0.187)	(0.824)	(0.326)	(1.508)
log(Mean Age)			-0.541	-3.446
			(1.645)	(7.512)
Fraction Interstate Migrants			0.302	1.542
			(0.240)	(1.123)
log(Value Manuf. Output)			0.067*	0.410
			(0.035)	(0.231)
log(Value Farm Product)			-0.083	-0.341
			(0.075)	(0.337)
# Counties	172	172	82	82
# Experiments	59	59	58	58
Adj. R-Sqr.	-0.007		-0.014	
F-Stat	0.619		0.839	
F-Test p-Value	0.604		0.558	
LR Chi-Sqr. Stat		1.981		6.745
LR-Test p-Value		0.576		0.456

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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

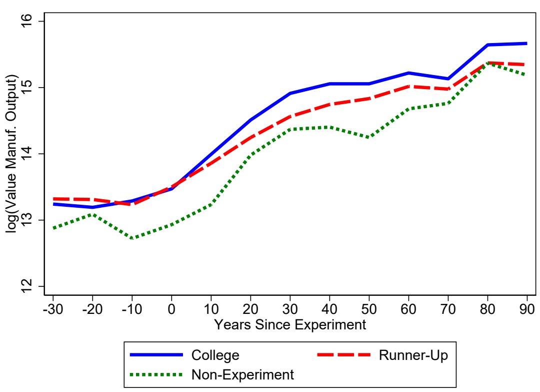
Figure A3: Time Series for Demographic and Economic Variables



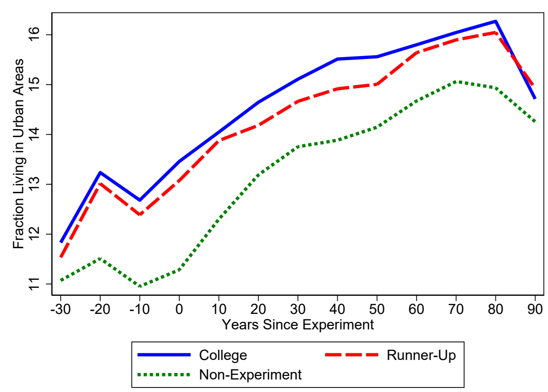
(a) log(Total Pop)



(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.

C 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 (Andrews, 2021*b*). 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.[5] 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 (2016*a*); see Petralia, Balland and Rigby (2016*b*) for details on the construction of this data.[6] Finally, for the years 1975 to 2010, contemporary digitized patent data sources can be used. I utilize the data available from the USPTO's PatentsView.[7] 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.[8] 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

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

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

[7] Available at `https://www.patentsview.org/download/`.

[8] The Jim Shaw, Annual Reports, and PatentsView data report the town and state of each inventor; the HistPat data reports the county and state of each inventor.

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.[9]

To obtain additional patent-level information, I merge by patent number from the Jim Shaw, Hist Pat, and PatentsView 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 et al., 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

[9]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.

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. In Section E.D I present results when using town-level or commuting zone-level data. Results are qualitatively similar, with any exceptions described in detail below.

Determining the county of each patent is non-trivial because each patent lists the town and state of each inventor, but not the county.[10] 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

[10]Except for the HistPat data, as noted above.

destroyed by fire, so I match town names to the 1900 census for the years in the 1890s. The results are insensitive to matching to the closest census rather than the previous census. 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 (2021*b*) 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.

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.

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

D.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 A10 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 A10: 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	26558	3313	15473
4	Georgia Institute of Technology	17	1917	1940	4309	0	1468
5	Iowa State University	31	1896	1940	8594	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 B.A. Table A11 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. Such a comparison is not possible for the vast majority of yearbooks (because the Commissioner of Education reports are only available for select years), 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 B.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, although counts of students and faculty tend to be clustered close to the 45-degree line.

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. Given the size of graduate programs for these years in the commissioner reports, and the fact that graduate students are likely highly geographically mobile, even extraordinarily high rates of patenting by graduate student alumni are unlikely to change the overall conclusions about the role of alumni in local patenting.

D.B Matching Patent and Yearbook Data to the Census

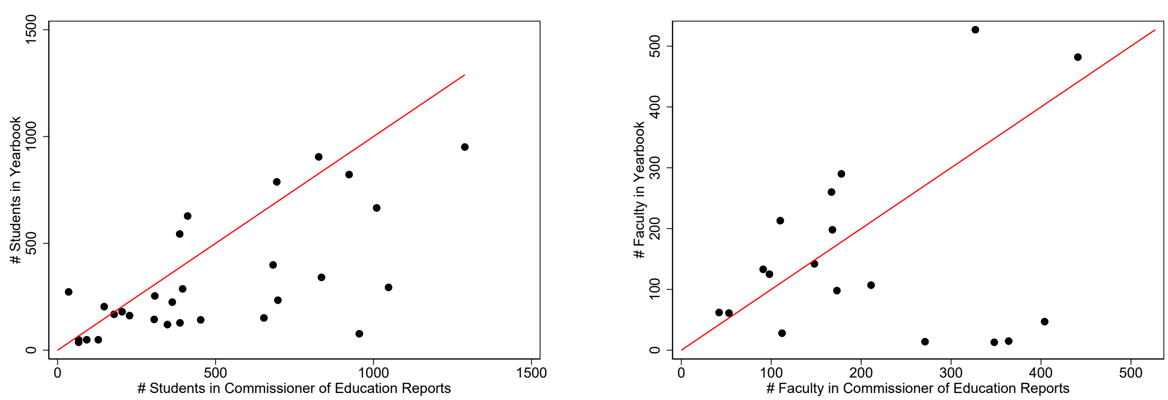
To determine the share of patents coming from alumni or faculty, I merge both the patent and yearbook data to the U.S. 100% decennial population census records, transcribed by `ancestry.com`, Family Search, and the Minnesota Population Center and hosted by the NBER. I proceed in eight steps.

Table A11: Comparing Yearbooks to Commissioner of Education Reports

	Yearbooks	Comm. Ed. Rep.	YB / Comm.Ed.
Num. Seniors	319.66	486.82	0.657
	(274.96)	(346.57)	(49.614)
	[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.

1. I prepare the census data for each census from 1850, 1860, 1870, 1880, 1900, 1920, 1930, and 1940. I restrict attention to males.[11] 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.

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.[12]

[11] 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.

[12] For examinations of inventor ages prior to 1940, see Sarada, Andrews and Ziebarth (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).

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, Andrews and Ziebarth (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.

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.[13] I then increase the size of the alumni

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

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.[14] That is, I initially calculate the share of patents belonging to each group without using the interpolated alumni counts; I discuss this 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

[14] 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.

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.[15]

D.C Match Rates

Table A12 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

[15] 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.

match 22% of the patents to the census in the full sample and 19% in the yearbook sample. These match rates are roughly double those in Sarada, Andrews and Ziebarth (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 F.A I show the sensitivity of results to using a match rate closer to that used in Sarada, Andrews and Ziebarth (2019). Second, in the matching procedure Sarada, Andrews and Ziebarth (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 patents to males in the census, whereas Sarada, Andrews and Ziebarth (2019) match to any gender. This will decrease my match rate relative to that in Sarada, Andrews and Ziebarth (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, Andrews and Ziebarth (2019) but match only select counties; the difference in sample could explain any further discrepancies.

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 A12: Match Rates

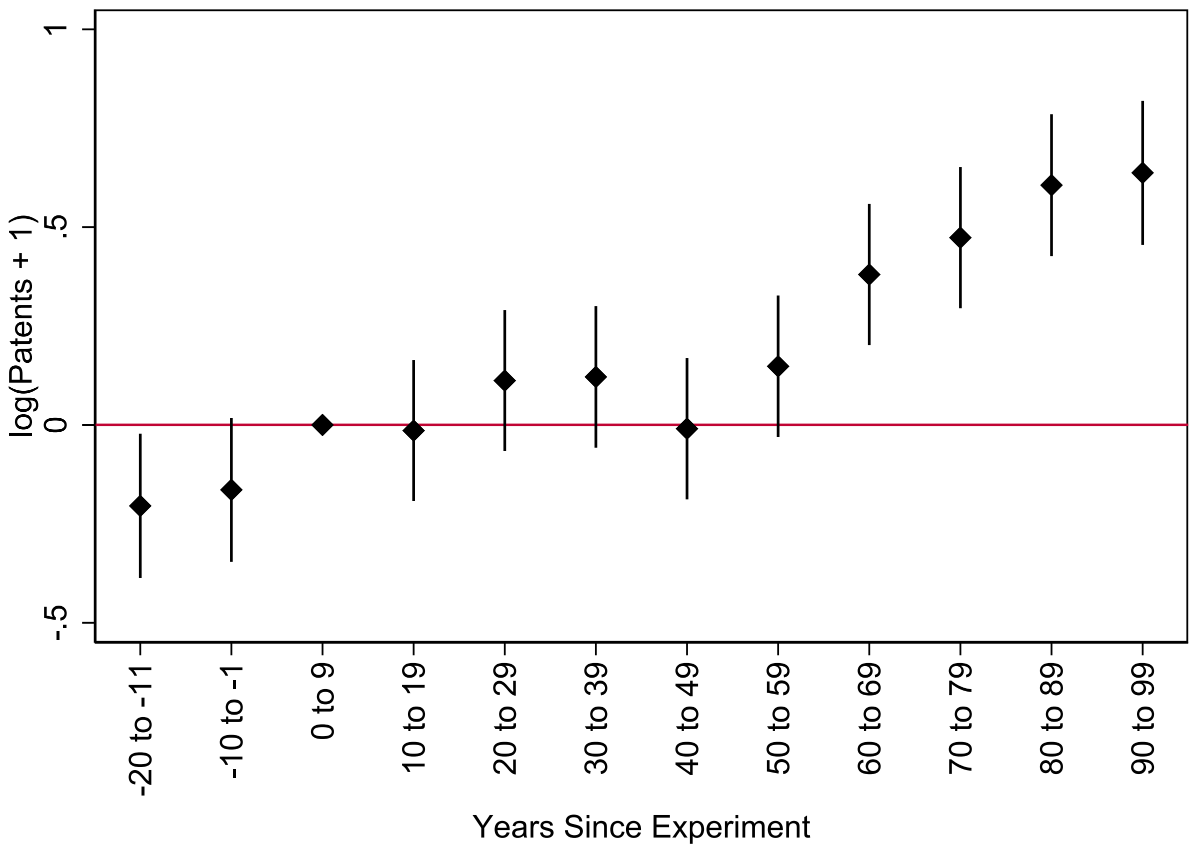
	Match Rates to Census
Patents	0.224
Patents (Yearbook Colleges)	0.189
Yearbooks (All)	0.042

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.

D.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 A13, 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. Figure A5 replicates Figure 4 using just the yearbook sample, and Figure A6 replicates Figure 5 using the yearbook sample.

Figure A5: Dynamics of the Effect of Local Colleges on Patenting with the Yearbook Colleges Sample



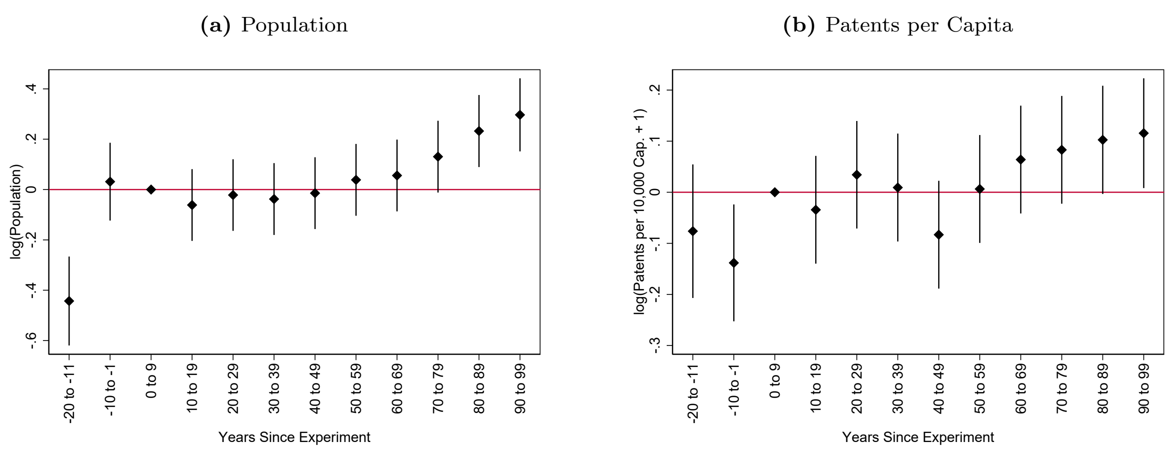
Notes: Estimated coefficients of the shift in logged patenting after establishment of a new college with a separate interaction term estimated for each time bin, along with 95% confidence bands. Time bins are are dummy variables that are equal to one for college counties in every ten year period before and after the establishment of the new college. The black diamonds show coefficients comparing the college counties to runner-up counties. Data are for the sample of colleges for which yearbook data are available.

Table A13: Baseline Regression Results with the Yearbook Colleges Sample

	log(Patents +1)	arcsinh(Patents)	Poisson	Any Patents
College * PostCollege	0.785***	0.922***	2.927***	0.205***
	(0.163)	(0.186)	(1.109)	(0.054)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	10,500	10,500	10,500	10,500
# Counties	60	60	60	60
# Experiments	20	20	20	20
Adjusted R-Squared	0.639	0.639		0.432
Log-Likelihood			-106,480.634	

Notes: Column 1 estimates the effect of establishing a college on local patenting when the dependent variable is $\log(Num.Patents + 1)$. The dependent variable in Column 2 is the inverse hyperbolic sine of patents. Column 3 presents results for a Poisson regression. Column 4 presents results of an extensive margin regression in which the dependent variable is an indicator equal to one if a county has at least one patent. Results are for the sample of colleges for which yearbook data are available. Standard errors are clustered by county and shown in parentheses.Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure A6: Dynamics of the Effect of Local Colleges on Population with the Yearbook Colleges Sample



Notes: Estimated coefficient of the shift in logged population (panel a) and patents per capita (panel b) 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 every ten years before and after the establishment of the new college. The black diamonds show coefficients comparing the college counties to runner-up counties. Data are for high quality experiments only. Data are for the sample of colleges for which yearbook data are available.

E Robustness Checks and Extensions for Baseline Results

E.A Runner-Up Counties that Receive a College at a Later Date

One question 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.[16] 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).[17]

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

[16] 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.

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

local effect of establishing a college.

To get a sense of the extent to which this issue can affect estimated magnitudes, in Column 1 of Table A14, 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, 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, so only the most successful colleges are included in the samples that exclude runners-up that get a college.

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:

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

The difference between this specification and the baseline specification in Equation (1) is the inclusion of county-by-experiment fixed effects and the 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 relatively few cases in which the same runner-up county appears in multiple experiments.

Table A14: 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	0.639***	0.656***	0.703***	0.696***	0.334***
	(0.140)	(0.140)	(0.165)	(0.186)	(0.102)
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
County-Year Observations	12,950	12,600	10,675	8,575	40,775
# Counties	71	69	59	47	180
# Experiments	27	27	24	20	73
Adjusted R-Squared	0.523	0.528	0.524	0.517	0.610

Notes: Column 1 excludes all runner-up counties with a traditional college in 2010. Column 2 also excludes all runner-up counties with a professional school in 2010. Column 3 additionally excludes all runner-up counties with a trade school or for-profit college in 2010. Column 4 additionally excludes all runner-up counties with a community college in 2010. 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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

E.B 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 A15.

Perhaps the most natural robustness exercise is to include experiment-by-year fixed effects to the baseline model, removing any experiment-specific time varying factors. More specifically, I estimate:

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

Results, presented in Column 1, are very similar to the baseline estimates, with college counties having 50.6 log points more patents per year than the runner-up counties after the college is established.

In Column 2, I estimate the following:

$$log(Patents + 1)_{it} = \delta_1 College_i * PostCollege_{it} + \delta_2 PostCollege_{it} + County_i + Year_t + \epsilon_{it}. \tag{3}$$

This is identical to the baseline specification in Equation (1) except for the inclusion of the $PostCollege_{it}$ indicator, which is equal to one for all counties i in the same experiment and all years t after the experiment's college has been established. 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. $PostCollege_{it}$ controls for changes that occur within all counties affiliated with a given college experiment after that experiment occurs, similar to the experiment-by-year fixed effects in the previous section. While the $PostCollege_{it}$ indicator does not have a natural economic interpretation, it can provide some suggestive information. For instance, if the coefficient on $PostCollege_{it}$ were significantly negative and similar in magnitude to the coefficient on $College_i * PostCollege_{it}$, this suggests that any increase in patenting in college counties is coming from a reallocation of activity away from the runner-up counties. Column 2 shows that this is not the case; the coefficient on $PostCollege_{it}$ is close to zero in magnitude and not statistically significant, while the coefficient on $College_i * PostCollege_{it}$ is similar to the baseline estimates in Table 2.[18]

In Column 3, I control for county-specific linear pretrends. In this specification, college counties have 47.4 log points more patents per year relative to the runner-up counties after establishing the new college, which is also very close to the baseline estimate.

In Column 4, I estimate an "intensive margin" specification, the complement to the extensive margin results in Column 4 of Table 2. I keep only the county-year observations with at least one patent and use logged patenting as the dependent variable; in contrast to other specifications, I do not transform the number of patents using the $log(Patents + 1)$ transformation. The results are a bit smaller in magnitude than the baseline estimates, with college counties having 27.5 log points more patents per year relative to the runner-up

[18] In prior versions of this paper, the specification in Column 2 was the baseline specification used in Table 2.

counties after establishing the college; this coefficient is not statistically significant. The lack of statistical significance is not surprising, as about 45% of the county-year observations are lost when restricting attention to counties with a positive number of patents. See Table A4 above for more information on how many counties have positive numbers of patents.

In Column 5, I use an alternative approach to handling counties with few patents, bottom coding the count of patents at the 5th percentile of patents in the U.S. in each year. In most years, the 5th percentile counties by patenting have zero patents, so this bottom coding method will still result in counts of zero in some years. I therefore again use $\log(Patents+1)$ as the dependent variable as in the baseline specifications. In this specification, establishing a new college causes 48.1 log points more patents per year relative to the runner-up counties after establishing the college, again nearly identical to the baseline estimates. In Column 6, I use an alternative bottom coding procedure, in which I bottom code patent counts using the 5th percentile of non-zero patents in the U.S. in each year. I then use $\log(Patents)$, without adding an arbitrary constant, as the dependent variable. With this bottom coding method, I find that establishing a college causes about 54.5 log points more patents per year in the college counties relative to the runners-up and the estimate is statistically significant at the 1% level.

Column 7 uses another alternative method to count patents, following the approach proposed by Blundell, Griffith and Van Reenen (1995). Rather than adding a positive constant before taking the log of patents, this alternative method uses $\log(Patents_{it})$ as the dependent variable. Whenever $Patents_{it} = 0$, a dummy variable is set to one and $\log(0)$ is replaced with 0. In this specification, establishing a new college leads to 34.4 log points more patents per year in the college counties relative to the runners-up.

In Column 8, I use another construction of logged patenting that adds a different arbitrary constant to the counts of patents, $\log(Patents + 0.0001)$. These results are much larger than the baseline estimate. This is not surprising in light of the intensive and extensive margin results, since $\log(Patents + 0.0001)$ specification penalizes having zero patents more heavily than the baseline specification that uses $\log(Patents + 1)$ as the dependent variable.

Finally, Column 9 displays the results using the number of patents as the dependent variable in a simple linear specification. Establishing a new college causes about 8.2 additional patents per year in the college counties. In sum, while the exact magnitude varies, all specifications tell the same story: establishing a new college causes a sizable increase in local patenting.

E.C Preexisting Colleges

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 A16.

Table A15: Additional Regression Specifications

	log(Patents +1)	log(Patents + 1)	County-Specific Trends	log(Patents) (Intensive Margin)	log(Patents + 1) (Bottom Coded)	log(Patents) (Bottom Coded)	Alternative log(Patents)	log(Patents + 0.0001)	Number of Patents
College * PostCollege	0.506***	0.499***	0.474***	0.275	0.481***	0.545***	0.344**	1.477***	8.164***
	(0.110)	(0.169)	(0.150)	(0.277)	(0.154)	(0.176)	(0.138)	(0.356)	(2.793)
PostCollege		-0.051							
		(0.093)							
Zero Pat. Dummy							-1.258***		
							(0.039)		
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Experiment FE	No	No	No	No	No	No	No	No	No
County-Experiment FE	No	No	No	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Experiment-Year FE	Yes	No	No	No	No	No	No	No	No
Year Range	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	33,425	33,425	33,425	18,229	33,425	33,234	33,425	33,425	33,425
# Counties	174	174	174	174	174	174	174	174	174
# Experiments	63	63	63	63	63	63	63	63	63
Adjusted R-Squared	0.726	0.610	0.616	0.634	0.610	0.686	0.717	0.519	0.423

Notes: Column 1 includes experiment-by-year fixed effects when the dependent variable is $log(Patents + 1)$. Column 2 includes county-specific linear time trends. Column 3 estimates an intensive margin model, including only counties-year observations in which a county had at least one patent and using $log(Patents)$ as the dependent variable. Column 4 bottom-codes county patents at the 5% percentile of patenting in each year and uses $log(Patents + 1)$ as the dependent variable. Column 5 bottom-codes county patents at the 5% percentile of non-zero patenting in each year and uses $log(Patents)$ as the dependent variable. Column 6 uses $log(Patents)$ as the dependent variable but sets the dependent variable equal to zero for any county-year observations with zero patents and includes an indicator variable for these observations. Column 7 uses $log(Patents + 0.0001)$ as the dependent variable. The dependent variable in column 8 is the number of patents. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

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 institutions (Kinnear, 1972; Wallenstein, 1997).[19] When excluding these cases, establishing a college causes about 61 log points more patents per year in the college county relative to the runners-up, relative to a baseline increase of 48.1 log points.

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.[20] The estimate is a bit larger than that in Column 1 and still statistically significant, with college counties producing 68.1 log points more patents per year relative to the 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 B.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

[19]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.

[20]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.

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 and the estimate in Column 3.

In short, across all specifications, it does not appear that the presence or absence of preexisting colleges substantially alters the interpretation of the results.

Table A16: 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	0.610***	0.681***	0.444***	0.360**
	(0.194)	(0.194)	(0.158)	(0.162)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	24,325	12,600	31,500	26,250
# Counties	134	68	163	140
# Experiments	48	22	60	51
Adjusted R-Squared	0.602	0.670	0.607	0.620

Notes: Column 1 excludes all cases in which the presence of a preexisting college was a factor in the decision of where to locate the college. Column 2 includes only the cases in which the presence of a preexisting college was a factor in the decision of where to locate the college. Column 3 includes only the cases in which the college county did not have any preexisting colleges at the time of the college site selection experiment. Column 4 includes only the cases in which none of the college or runner-up counties had any preexisting colleges at the time of the college site selection experiment. 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.D 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 A17 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 when using commuting zones defined for 1990 or 2000. The estimated treatment effect is a bit larger than the baseline result, with commuting zones that receive a new college having about 57.1 log points more patents per year relative to their runner-up commuting zones after the college is established. While it is encouraging that this result is qualitatively similar to the baseline estimate, I prefer to use counties as the baseline level of geographic aggregation throughout the paper. Commuting zones were determined based on modern patterns of economic activity, and hence are endogenous to the historical establishment of colleges.

In Column 2 of Table A17 I present results at the town or municipality level. These results, using a smaller level of geographic aggregation, also have several drawbacks. First, some specifications include control variables that are only available at the county level. Second, and perhaps more important, the towns of residence listed on patents are often unreliable for historical patents. Inventors are free to record their location any way that they want, and especially outside of major cities town or municipal boundaries may not have been well-defined or inventors may have erroneously recorded their location of residence as the nearest town. Inside large cities, inventors may have recorded their location as a neighborhood rather than the city (e.g., Charlestown instead of Boston, MA). While there is thus substantial uncertainty about a patent's municipality, it is usually easy to assign a patent to a county; all neighborhoods within a city will be in the same county, as will typically outlying areas surrounding a town. Finally, municipal boundaries have changed much more over time than have county borders. I find that a town that receives a college has a statistically significant 8.8 log points more patents per year than the runner-up towns after establishing the college. This estimate is smaller than the county and commuting zone estimates, consistent with attenuation from the measurement errors described above as well as substantial geographic spillovers across town borders.

E.E 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.

E.F 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 from pre-treatment years. 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 A17, 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.

Table A17: Results at Other Geographic Levels and a Placebo Test

	Commuting Zones Zones	Towns	log(Patents + 1)	log(Population)
College * PostCollege	0.571***	0.088**	0.044	0.166
	(0.159)	(0.040)	(0.060)	(0.122)
County FE	No	No	Yes	Yes
Experiment FE	Yes	No	Yes	Yes
County-Experiment FE	No	No	Yes	Yes
Commuting Zone FE	Yes	No	No	No
Commuting Zone-Experiment FE	Yes	No	No	No
Town FE	No	Yes	No	No
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-1945	1836-1953	1840-1950
County-Year Observations	25,200	25,440	8,839	689
# Counties	136	234	191	173
# Experiments	73	73	63	62
Adjusted R-Squared	0.630	0.268	0.217	0.540

Notes: In Column 1, I present baseline regression results estimated at the commuting zone level rather than the county level. In Column 2, results are estimated at the town level rather than the county level. In Columns 3 and 4, 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 Columns 1-3 the dependent variable is $log(Patents+1)$, while in Column 4 it is $log(Population)$. Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

E.G 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 A18 I use the patent classes assigned to historical patents by Marco et al. (2015) to examine how patenting across all classes changes after establishing a new college.[21] 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 about 11% the size of the baseline estimate and is no longer statistically significant.

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

$$\begin{aligned} PatentMeasure_{ijct} = &\delta_1 College_i * PostCollege_{it} + \delta_2 PostCollege_{it} \\ &+ County_i + Class_c + Year_t + Class_c * Year_t + \epsilon_{ijct}, \end{aligned} \tag{4}$$

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 larger than that in Column 1 but still statistically insignificant and much smaller than the baseline estimates. Thus, in addition to changing the number of patents in a county, establishing a college appears to substantially shift the technology classes in which counties patent.

The results in Columns 1 and 2 suggest that shifting in the composition of types of inventions patented in college counties after a new college is established accounts for a large share of the increases in patenting. Are college counties becoming increasingly specialized

[21] 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.

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:

$$PatentConcentration_{it} = \sum_{c \in C_{it}} \left(\frac{Patents_c}{\sum_{k \in C_{it}} Patents_k} \right)^2 \tag{5}$$

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 A18. 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.

E.H 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

Table A18: Patent Classes

	Control for Patent Classes	By Patent Classes	Class Concentration	Class Concentration
College * PostCollege	0.063	0.146	587.970	-67.005
	(0.199)	(0.172)	(548.692)	(205.333)
Number of Patents				30.928***
				(5.643)
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	1844-1975	1836-2010	1836-2010
County-Year Observations	12,683	75,537	8,393	8,393
# Counties	174	174	174	174
# Experiments	63	63	63	63
Adjusted R-Squared	0.602	0.528	0.323	0.670

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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

distribution of inventive abilities, in which case the distribution of patent qualities may not change at all. Following Hall, Jaffe and Trajtenberg (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 A19, I show that the absolute number of patent citations in college counties increases by 69.4 log points relative to the runner-up counties after establishing a new college. This is 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) increases by about 2.5 citations per patent in college counties relative to the runners-up after the college is established, which is statistically significant at the 1% level. 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. In Column 4, I also control for the distribution of patent classes in each county as in Column 1 of Table A18; this is due to the fact that some classes may inherently receive more citations than others. The coefficients in Columns 3 and 4 are similar in magnitude and significance to that in Column 2.

To get a better sense of whether the increasing average patent quality is driven by more "superstar" patents or a rightward shift of the entire distribution, 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 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, patents in college counties appear to receive more citations, although it is less clear exactly how the distribution of citations shifts in college counties relative to the runners-up after establishing the college.

Unfortunately, patent citations are only consistently available beginning in 1947, making them a less-than-ideal measure when using historical patent data. This means that patents from the 19th century which may have been highly impactful for decades will still being recorded as having zero forward citations if they were not cited by another patent that issued after 1947. This is likely to include most of the patents in the years prior to college establishment. 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 typically the most broad. A very short first claim therefore indicates a patent with a very broad legal 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 A20, 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. Thus when using this alternative measure of patent quality, which may be more informative over the entire length of the sample period, establishing a college has no measurable effect on patent quality.

Table A19: Patent Quality: Forward Citations

	log(Citations + 1)	Citations per Patent	Citations per Patent	Citations per Patent	Fraction of Citations below 10th Percentile	Fraction of Citations above 90th Percentile
College * PostCollege	0.694***	2.524***	2.680***	2.644***	5.455	7.724
	(0.204)	(0.898)	(0.514)	(0.514)	(4.898)	(7.248)
Zero Pat. Dummy			-3.555***	-3.942***		
			(0.309)	(0.403)		
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
County-Year Observations	33,288	18,229	33,425	33,425	5,981	5,981
# Counties	174	174	174	174	173	173
# Experiments	63	63	63	63	63	63
Adjusted R-Squared	0.715	0.423	0.412	0.413	0.506	0.485

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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

E.I Results with Low Quality Site Selection Experiments

In Table A21, I repeat the analysis in 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 were nevertheless not as good as randomly assigned and so are excluded from the baseline sample. Instead of

Table A20: Patent Quality: Claim Length

	Length of 1st Claim	log(Length of 1st Claim)	Fraction of 1st Claim below 10th Percentile	Fraction of 1st Claim above 90th Percentile
College * PostCollege	-1.077	-0.035	-0.002	0.008
	(2.992)	(0.025)	(0.016)	(0.018)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	5,687	5,687	5,687	5,687
# Counties	173	173	173	173
# Experiments	63	63	63	63
Adjusted R-Squared	0.336	0.386	0.022	0.022

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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

estimating Equation (1), I now estimate a triple-difference equation of the form

$$\begin{aligned} PatentMeasure_{ijt} = & \delta_1 College_{ij} * HighQuality_j * PostCollege_{ijt} \\ & + \delta_2 College_{ij} * PostCollege_{ijt} + \delta_3 HighQuality_j * PostCollege_{ijt} \\ & + \delta_4 PostCollege_{ijt} + County_i + Year_t + Experiment_j + \epsilon_{it}. \end{aligned} \tag{6}$$

The indices mean the same as in the previous equations; I also index experiments by j as in Equation (2) since with the larger sample of colleges it may be the case that a county is a runner-up county in multiple experiments. Now $HighQuality_j$ is equal to one if experiment j is included in the original baseline sample (that consists of only the high quality experiments) and zero otherwise.

I estimate Equation (6) using the same dependent variables as in Columns 1-4 of Table 2. In the new regression specifications, the coefficient of the triple-interaction term, $College_{ij} * HighQuality_j * PostCollege_{ijt}$, 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, although not statistically significant, when using logged patenting and the inverse hyperbolic sine of patenting as the dependent variables, negative and statistically significant in the Poisson regression, and positive and statistically significant at the 10% level in the extensive margin model. In most specifications, it therefore appears that the low quality experiments over-estimate the effect of a college relative to the high quality experiments, although this difference is typically imprecisely estimated. The coefficient on $College_{ij} * PostCollege_{ijt}$ 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 magnitudes are larger for all specifications except that in Column 4. 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.[22] Combining these coefficients reveals that high quality college counties increase patenting by amounts broadly similar to the findings in Columns 1-4 of Table 2.

In Figure A7, I present the analog to Figure 4 but compare only the low quality colleges to their runner-up counties. More precisely, I estimate

$$
\begin{aligned}
PatentMeasure_{ijt} = & \sum_{\tau \in T} [\beta_{1\tau} College_{ij} * TimeBin_{\tau} + \delta_{2\tau} TimeBin_{\tau}] \\
& + County_i + Year_t + Experiment_j + \epsilon_{ijt}. \qquad (7)
\end{aligned}
$$

for the low quality colleges j and plot the $\beta_{1\tau}$ coefficients. While the overall dynamics are similar to those for the baseline sample in Figure 4, a pre-trend is apparent, confirming the suspicion that runner-up counties are less suitable as counterfactuals for the low quality experiments.

E.J 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 A22, I reproduce the baseline result on patenting using only

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

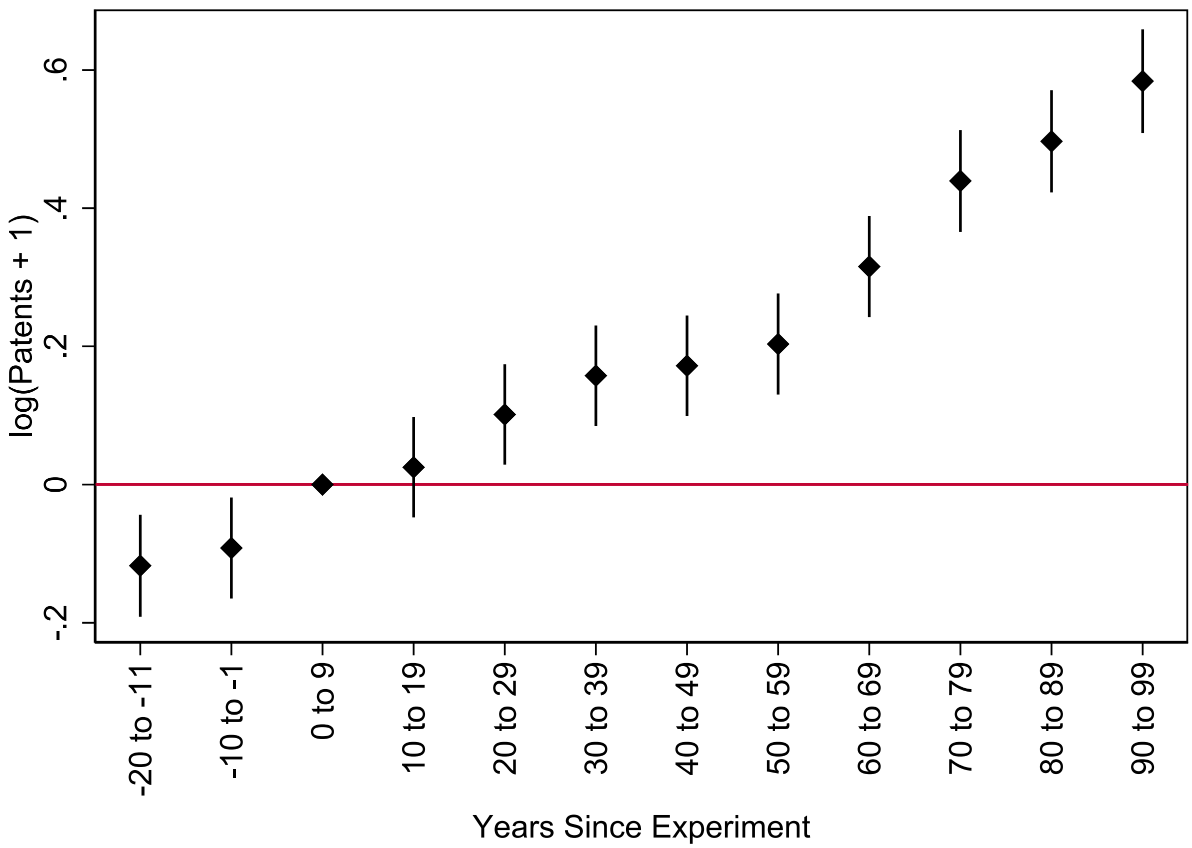
$$
\begin{aligned}
(E[y_{College,HighQuality,PostCollege}] - & E[y_{College,HighQuality,PreCollege}]) \\
& - (E[y_{RunUp,HighQuality,PostCollege}] - E[y_{RunUp,HighQuality,PreCollege}]) \\
= & [\delta_1 + \delta_2 + \delta_3 + \delta_4] - [0] - [\delta_3 + \delta_4] + [0] \\
= & \delta_1 + \delta_2.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

Table A21: Results with High and Low Quality College Site Selection Experiments

	log(Patents +1)	arcsinh(Patents)	Poisson	Any Patents
College * HighQuality * PostCollege	-0.238	-0.226	-0.281**	0.065*
	(0.198)	(0.216)	(0.136)	(0.034)
College * PostCollege	0.656***	0.725***	1.939***	0.072***
	(0.148)	(0.161)	(0.372)	(0.024)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Experiment FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	133,845	133,845	133,845	133,845
# Counties	462	462	462	462
# Experiments	181	181	181	181
Adjusted R-Squared	0.662	0.664		0.465
Log-Likelihood			-3,647,787.791	

Notes: Column 1 estimates the effect of establishing a college on local patenting when the dependent variable is $\log(Patents + 1)$. The dependent variable in Column 2 is the inverse hyperbolic sine of patents. Column 3 presents results for a Poisson regression. Column 4 presents results of an extensive margin regression in which the dependent variable is an indicator equal to one if a county has at least one patent. All columns use both high and low quality 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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure A7: Dynamics of Treatment Effect for the Low Quality Experiments



Notes: Estimated coefficients of the shift in logged patenting in college counties 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. The black diamonds show coefficients comparing the college counties to runner-up counties. Data are for low quality college site selection experiments.

patenting in census years. The estimated coefficient is similar to the baseline coefficient estimated in Column 1 of Table 2. Column 2 reproduces the results from Column 1 of Table 4 and shows the effect of a new college on logged county population.

In Column 3, I re-estimate Equation (1) but include $\log(Population)$ as a control. Not surprisingly, county population is highly predictive of county patenting (a ten percent increase in population leads to a 7.8% increase in patenting). When including $\log(Population)$, the coefficient on the interaction term of interest is only about 15.8% of the baseline estimate, decreasing from 48.4 log points more patents per year in the baseline to 0.9 log points 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 13.4% of its baseline value. Thus, simply controlling for population in the baseline regression can explain about 85% 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 since it is also affected by the treatment of establishing a college.

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

$$\begin{aligned}\log(Patents_{it}+1) =& \delta_1 College_i * PostCollege_{it} + \delta_3 \log(Population_{it}) \\ &+ \delta_3 College_i * PostCollege_{it} * \log(Population_{it}) \\ &+ County_i + Experiment_i + Year_t + \epsilon_{it}. \end{aligned} \tag{8}$$

Results are presented in Column 5 of Table A22. There is some evidence that increasing population increases patenting more in college counties (measured by δ_3). A 10% increase in population increases the differences-in-differences estimate by a statistically significant 5.9%. Results are similar when using other functional forms or semiparametric regressions for county population.

Appendix G.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 A22: The Effect of Colleges on Patenting when Controlling for Population

	log(Patents + 1)	log(Population)	log(Patents + 1)	log(Patents + 1)	log(Patents + 1)
College*PostCollege	0.484**	0.503***	0.094	0.080	-6.178***
	(0.190)	(0.169)	(0.121)	(0.090)	(0.718)
log(Population)			0.775***		0.665***
			(0.071)		(0.076)
College * PostCollege * log(Total Pop.)					0.592***
					(0.068)
Population Fract. Polynomials	No	No	No	Yes	No
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1840-2010	1840-2010	1840-2010	1840-2010	1840-2010
County-Year Observations	3,230	3,230	3,230	3,230	3,230
# Counties	174	174	174	174	174
# Experiments	63	63	63	63	63
Adjusted R-Squared	0.692	0.723	0.823	0.887	0.847

Notes: Column 1 estimates the effect of establishing a college on local patenting when the dependent variable is $\log(Patents + 1)$. Column 2 estimates the effect of establishing a college on local population when the dependent variable is $\log(Population)$. The dependent variable for Columns 3-5 is $log(Patents+1)$. Column 3 re-estimates Column 1 but includes a control for $log(Population)$. Column 4 re-estimates Column 1 but includes fractional polynomial controls for population. Column 5 estimates the effect of establishing a college on local patenting when the dependent variable is $\log(Patents + 1)$ when controlling for $log(Population)$ and interacting $log(Population)$ with a dummy for college counties, a dummy for post-college years, and the interaction term. Results are from census years only and $\log(Patents + 1)$ 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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

F Additional Results on Patenting by Alumni and Faculty

F.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 assumptions.

In the baseline results, shown in Table 5, 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 (Bailey et al., 2020). In Table A23, 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 11.7% of all patents instead of 11.9% in the baseline results.

In the baseline results, I consider two records to be a match if they have a bigram matching score of 0.8 or above.[23] In Table A24, 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

[23]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.

strictness reduces the share of patents attributed to alumni and faculty by more than 85%; they now account for only about 1.6% of patents. I am still able to match a similar number of patents to the census with this new criteria. A ratio of 0.85 is still quite liberal as a match cutoff; for instance, Sarada, Andrews and Ziebarth (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.

One of the concerns with using yearbook data, described in detail above in Section D, 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 A25, I present results without making this adjustment for missing yearbook years. Not surprisingly, without this adjustment alumni and faculty account for only about 5.5% of the patents in their college counties. Table A26 presents results when using an alternative method to interpolate student and faculty counts for missing yearbook years. In Table A26, I use cubic splines to interpolate counts, whereas for the baseline results in Table 5 I use a linear interpolation. Results are similar in both tables.

Finally, I conduct additional analysis to test the robustness of the conclusion that "post-college others" account for more than twice the share of patents as do the "pre-college others." 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 A27, 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 relaxed matching criteria results in a much larger share of "pre-college others," since any patentee with a common name is likely to appear in multiple censuses; now "pre-college others" account for 69% of patents, while "post-college others" account for 19%.

As described in detail in Section D.B above, an individual is considered to be a "pre-college other" if the individual is not an alumnus or faculty and was present in the college's county in the last decennial census prior to establishing the college. The problem is that it is impossible to know when individuals migrate in between census years. For instance, an individual who migrated to a college county in 1871 would not appear in that county in the 1870 census, but would reside in the county prior to the college if the college was established in 1872 or later. Thus, the decennial nature of the census records may induce measurement error. In Table A28, I test for whether this source of bias is likely to be quantitatively meaningful by regressing the share of "post-college others" on how many years after the previous census each college was established.[24] Column 1 shows that each additional year further away from the census in which a college was established increases the share of post-college others in the entire population by about 1.1 percentage points. While each additional year is relatively small in magnitude and statistically insignificant, a college that is established nine years later sees 9.9 percentage points more post-college others, about 12.6% of the average share of post-college others. In Column 2, I repeat this exercise but calculate the share of post-college others among patentees. The correlation is similar in

[24]So, a college established in 1879 was established nine years after the 1870 census. Note that there was no census in 1890, so for colleges established in the 1890s I calculate the number of years to the 1880 census.

magnitude and still statistically insignificant. In Columns 3 and 4 I repeat these exercises but use the naive name matching procedure as in Table A27. The correlation for the share of post-college others in the entire population and among patentees is about one-third the magnitude of the original estimate, although now one is positive and one negative; again, neither is statistically significant. In sum, the sign of these correlations are sensitive to the specification and are statistically insignificant. Nevertheless, in some specifications, a college established almost a decade after the prior census may have a non-trivially larger share of post-college others.

To further address this issue, I repeat the matching results using an alternative definition of "pre-college" and "post-college others." I redefine "pre-college others" to be any individual who was located in a census in the college county through the first census *after* a college was established. This is therefore an upper bound on the number of pre-college others, since it includes some individuals who migrated to the college county in between the establishment of the college and the next decennial census. Results are presented in Table A29. While the share of pre-college others is a bit larger than the baseline, post-college others still account for a plurality of population and patents (53.2% of population and 48.8% of patents, compared to 41.2% and 39.3% of population and patents, respectively, for the pre-college others). Thus, measurement error induced by the decennial nature of the census data is not driving the main qualitative conclusions. This should perhaps not be surprising, since most of the post-college others likely arrive long after the college is established.

As a final test to ensure that I am not inflating the counts of post-college others, I exclude from the analysis any of the colleges that were established prior to 1860. For these early colleges, there may not be any usable census data in the "pre" period, and so all non-alumni and non-faculty inventors will be classified as post-college others. I present the results in Table A30. In spite of the fact that, by construction, this sample includes only younger colleges that have had less time to grow and build up a stock of alumni, the shares for all four groups, including the pre- and post-college others, are nearly identical to the baseline results.

Table A23: Patents by Alumni, Faculty, and Others: Multiple Matches

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	277.626	0.055	0.079	0.116
	(526.578)	(0.070)	(0.246)	(0.197)
Faculty	4.115	0.001	0.000	0.001
	(6.375)	(0.001)	(0.003)	(0.003)
Pre-College Others	1,151.105	0.227	0.186	0.266
	(2,317.772)	(0.190)	(0.809)	(0.327)
Post-College Others	3,640.316	0.718	0.434	0.617
	(7,274.108)	(0.222)	(1.285)	(0.354)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

F.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

Table A24: Patents by Alumni, Faculty, and Others: Strict Matching Criteria

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	277.183	0.055	0.011	0.016
	(525.497)	(0.070)	(0.059)	(0.046)
Faculty	4.115	0.001	0.000	0.000
	(6.375)	(0.001)	(0.000)	(0.000)
Pre-College Others	1,151.268	0.227	0.197	0.296
	(2,318.205)	(0.190)	(0.847)	(0.359)
Post-College Others	3,640.496	0.718	0.460	0.689
	(7,274.211)	(0.222)	(1.342)	(0.350)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A25: Patents by Alumni, Faculty, and Others: No Adjustment for Missing Yearbooks

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	48.392	0.010	0.037	0.053
	(140.987)	(0.033)	(0.243)	(0.178)
Faculty	1.512	0.000	0.001	0.002
	(5.330)	(0.001)	(0.036)	(0.030)
Pre-College Others	1,232.474	0.243	0.198	0.281
	(2,517.697)	(0.207)	(0.848)	(0.345)
Post-College Others	3,790.784	0.747	0.469	0.665
	(7,412.226)	(0.213)	(1.368)	(0.373)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A26: Patents by Alumni, Faculty, and Others: Alternative Between-Yearbook Interpolation

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	282.680	0.056	0.080	0.118
	(525.982)	(0.069)	(0.247)	(0.197)
Faculty	7.787	0.002	0.001	0.001
	(7.930)	(0.004)	(0.006)	(0.007)
Pre-College Others	1,149.591	0.227	0.185	0.265
	(2,318.619)	(0.189)	(0.808)	(0.326)
Post-College Others	3,633.103	0.716	0.433	0.616
	(7,273.377)	(0.221)	(1.285)	(0.353)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A27: Patents by Alumni, Faculty, and Others: Naive Matching Criteria

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	277.626	0.055	0.080	0.118
	(526.578)	(0.070)	(0.247)	(0.197)
Faculty	4.115	0.001	0.001	0.001
	(6.375)	(0.001)	(0.006)	(0.007)
Pre-College Others	3,382.254	0.667	0.487	0.694
	(6,211.392)	(0.285)	(1.557)	(0.353)
Post-College Others	1,409.167	0.278	0.131	0.186
	(3,216.698)	(0.297)	(0.604)	(0.337)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A28: Correlation between Share of Post-College Others and Years to the Prior Census

	Share Post-College	Share Post-College Patents	Share Post-College	Share Post-College Patents
Years to Census	0.011	0.012	0.003	-0.004
	(0.012)	(0.023)	(0.022)	(0.026)
Experiment FE	No	No	No	No
Year FE	No	No	No	No
Year Range	1839-1924	1839-1924	1839-1924	1839-1924
# Experiments	20	20	20	20
Mean of Dep. Var.	0.782	0.778	0.298	0.234
Adjusted R-Squared	-0.004	-0.047	-0.057	-0.065

Notes: Correlations between the baseline share of post-college others and the number of years between the last census prior to college establishment and the establishment of the college. The first column shows the correlation between the share of the entire county population that is post-college others and the years to the last census. The second column shows the correlation between the share of patentees that is post-college others and the years to the last census. The third and fourth columns repeat Columns 1 and 2, respectively, but use the naive matching criteria from Table A27 to determine matches. Standard errors are clustered by county and shown in parentheses.

Table A29: Patents by Alumni, Faculty, and Others: Alternative Definition of Pre- and Post-College

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	277.626	0.055	0.080	0.118
	(526.578)	(0.070)	(0.247)	(0.196)
Faculty	4.115	0.001	0.001	0.001
	(6.375)	(0.001)	(0.006)	(0.007)
Pre-College Others	2,090.429	0.412	0.275	0.393
	(3,593.553)	(0.178)	(0.996)	(0.369)
Post-College Others	2,700.993	0.532	0.343	0.488
	(5,575.431)	(0.207)	(1.169)	(0.377)

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county when using an alternative definition of pre- and post-college others in which individuals are counted as pre-college others if they reside in the college county in the first census after the college is established. 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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

Table A30: Patents by Alumni, Faculty, and Others: Excluding the Early College Site Selection Experiments

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	291.079	0.055	0.084	0.118
	(537.382)	(0.070)	(0.253)	(0.197)
Faculty	4.336	0.001	0.001	0.001
	(6.472)	(0.001)	(0.006)	(0.007)
Pre-College Others	1,213.411	0.229	0.195	0.266
	(2,363.795)	(0.189)	(0.829)	(0.326)
Post-College Others	3,780.515	0.715	0.455	0.614
	(7,442.694)	(0.221)	(1.315)	(0.353)

Notes: Population and patenting results for college alumni, faculty, and other individuals living in the same county when excluding any colleges established before 1860. 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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

other local colleges could be contributing a large number of local patents.

To check for this, I repeat the results in Table 5 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 A31. When restricting attention to these colleges, alumni and faculty together account for about 1.5% of all patents, far smaller than the 11.9% 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.

Table A31: Patents by Alumni, Faculty, and Others: No Counties with Multiple Colleges

	Number of People	Share of Population	Number of Patents	Share of Patents
Alumni	92.365	0.044	0.002	0.015
	(135.952)	(0.075)	(0.016)	(0.094)
Faculty	2.597	0.001	0.000	0.000
	(2.916)	(0.002)	(0.000)	(0.000)
Pre-College Others	303.256	0.143	0.048	0.200
	(489.401)	(0.171)	(0.378)	(0.382)
Post-College Others	1,715.185	0.812	0.190	0.786
	(1,961.039)	(0.181)	(0.727)	(0.387)

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. Results are for college counties for which yearbook data is available. Standard deviations are shown in parentheses.

F.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 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.

F.C.1 Alumni Patenting

To construct measures of alumni patenting in post-1940 years, I use recent work by Bell et al. (2019), who link patents to both IRS tax records and to alumni records used in Chetty et al. (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.[25] To make the Bell et al. (2019) data consistent with the results from the college yearbooks, I restrict attention to only the colleges for which yearbook data is available.

As when using the yearbook data, the Bell et al. (2019) data 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.[26] 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 D.A for the historical yearbook data.

One important difference between these alumni patenting counts and the pre-1940 counts

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

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

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 et al. (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.[27]

With all these adjustments in mind, the alumni account for about 13.1% 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 5. Even under liberal alternative assumptions, alumni still account for less than a quarter of all patents in college counties, and likely much less.

F.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

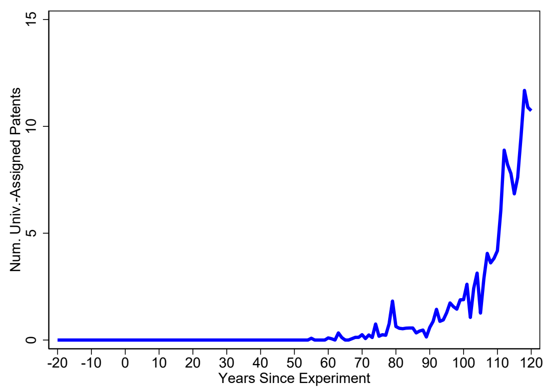
[27] I also calculate results when scaling the number of alumni patents by 0.22, which is the share of alumni living in the same state as their college according to Zolas et al. (2015). This should be thought of as an upper bound on the share of alumni remaining in the county of their alma mater. Under this alternative assumption, alumni still account for only 23% of all patents in the college county.

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 faculty and would be unlikely to be listed in a college yearbook. Note that university-assigned patents may also include some patents by alumni if the alumni worked on their invention while they were students or assigned their patents to their alma maters for any reason.

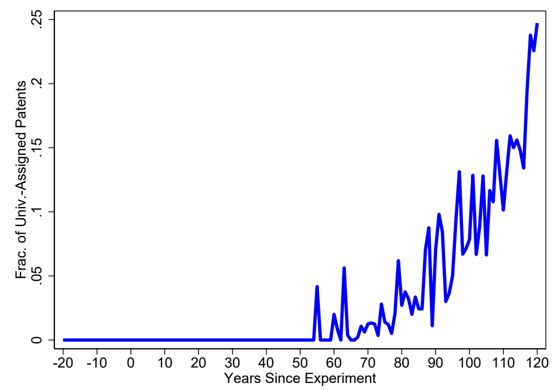
To create a list of university-assigned patents, for every patent granted in a college county, I check the assignee name for the words "College," "University," "Institute," or any of their common abbreviations. Note that this will capture *all* university-assigned patents in a county, and not just the 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.

Results are presented in Figure A8. 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 a few outlier years, 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 A8: University-Assigned Patents



(a) Number of 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.

G Additional Results Investigating the Indirect Channels

G.A Most Common Inventor Occupations by Decade

Table A32 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, Andrews and Ziebarth (2019). I use the variable "occ1950" to get consistent occupation names across censuses.

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.[28]

Table A32: Most Common Inventor Occupations by Year

	1900 Occ.	Share	1910 Occ.	Share	1920 Occ.	Share	1930 Occ.	Share	1940 Occ.	Share
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, Andrews and Ziebarth (2019), along with the percentage of inventors belonging to each occupation code in each year.

G.B Consolation Prizes

G.B.1 Additional Details on the Consolation Prizes

In this section, I present additional details on the consolation prize counties. First, I supplement the results in Figures 6 and 7 by showing that not only were college counties similar to their consolation prize counties in the last census before the colleges were established, but they were evolving similarly as well. Results are presented in Figure A9, showing the evolution of logged population, the fraction of the population living in urban areas, logged farm product, and logged manufacturing output. This figure is constructed similarly to Figure A3.

Next, I provide additional information on the consolation prize sites. While Figures 6 and 7 show that the counties that received consolation prizes were similar to the college

[28]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.

counties in the year before the colleges and consolation prizes were established, it is possible that the surrounding areas may have been different between the college and consolation prize counties. In particular, one might wonder if the fact that consolation prize counties patent similarly to the college counties even after the college is established—which I highlight in Table 8—is driven by the fact that consolation prizes happen to be built close to large population centers or other non-consolation prize-related features that could plausibly explain their patenting, while colleges are built in more remote locations. The nature of the college site selection experiments ensures that, on average, college and runner-up sites should be in places with similar characteristics, including similar remoteness and distance to major cities. But since there is only a small number of consolation prize cases, it is possible that the consolation prize counties are substantially less remote than their college county counterparts.

I present information on the location of the consolation prizes in Table A33. For each consolation prize, I list the distance to the nearest "major" city, as well as the distance between the consolation prize and that major city in miles and the current drive time according to Google Maps (calculated at a time with minimal traffic).[29] The college and consolation prizes are typically similarly remote. Furthermore, in most cases, the consolation prize counties are quite distant from any major cities, more than an hour's drive in most cases and more than two hours in five of the 13 cases. As travel times have trended downward through history, the time to reach a major city from these consolation prize locations would have

[29]The decision of what counts as a major city was somewhat subjective. I do not count any of the college or consolation prize counties as sites of major cities. For instance, the closest major city to Stutsman or Burleigh Counties in North Dakota would be Fargo, the site of North Dakota State University, or Grand Forks, the site of the University of North Dakota. Bismarck, ND, is the site of the consolation prize, the state capital, and currently the second largest city in North Dakota. I instead consider the nearest major city for Stutsman and Burleigh Counties to be Minneapolis, MN.

been longer than the listed times for most of the sample period.

There are three exceptions in which the consolation prize is less than 50 miles from a major city: Weber County, UT, is 43 miles from Salt Lake City; Canon City in Fremont County, CO, is 45 miles from Colorado Springs; and Salem in Marion County, OR, is 46 miles from Portland. While none of these is among the largest cities in the country, all three are substantially larger than the college county. While Salt Lake City is today a major innovation hub, this is a relatively recent development (Vara, 2015), so it is not clear that proximity to the city would have been a major boon to Weber County for most of its existence. Colorado Springs likewise did not become an important hub until the Air Force Academy located there in 1954 (Fagan, 2006; Nauman, 2004). Moreover Canon City is further from Colorado Springs than Boulder (the site of the University of Colorado) is from Denver, the largest city in the state; if anything Boulder's relative proximity to Denver might be expected to overstate the effect of the college relative to the consolation prize. Portland, OR, on the other hand, was among the 50 largest cities in the U.S. for most of the years after Oregon State was established, raising the possibility that Portland's success explains the high levels of patenting at the state capital in Salem relative to Corvallis, the site of Oregon State University. Given that all three cities are more than 40 miles from the consolation prize sites, I find this explanation unlikely, especially in years before widespread suburbanization. In Section G.B.3 below, I show that college and consolation prize counties still patent similarly even when excluding these three cases, as well as when excluding recent decades during which suburbanization may have linked labor markets in consolation prize counties with those of major cities.

Table A34 lists contemporary outcomes for the college and consolation prize counties.

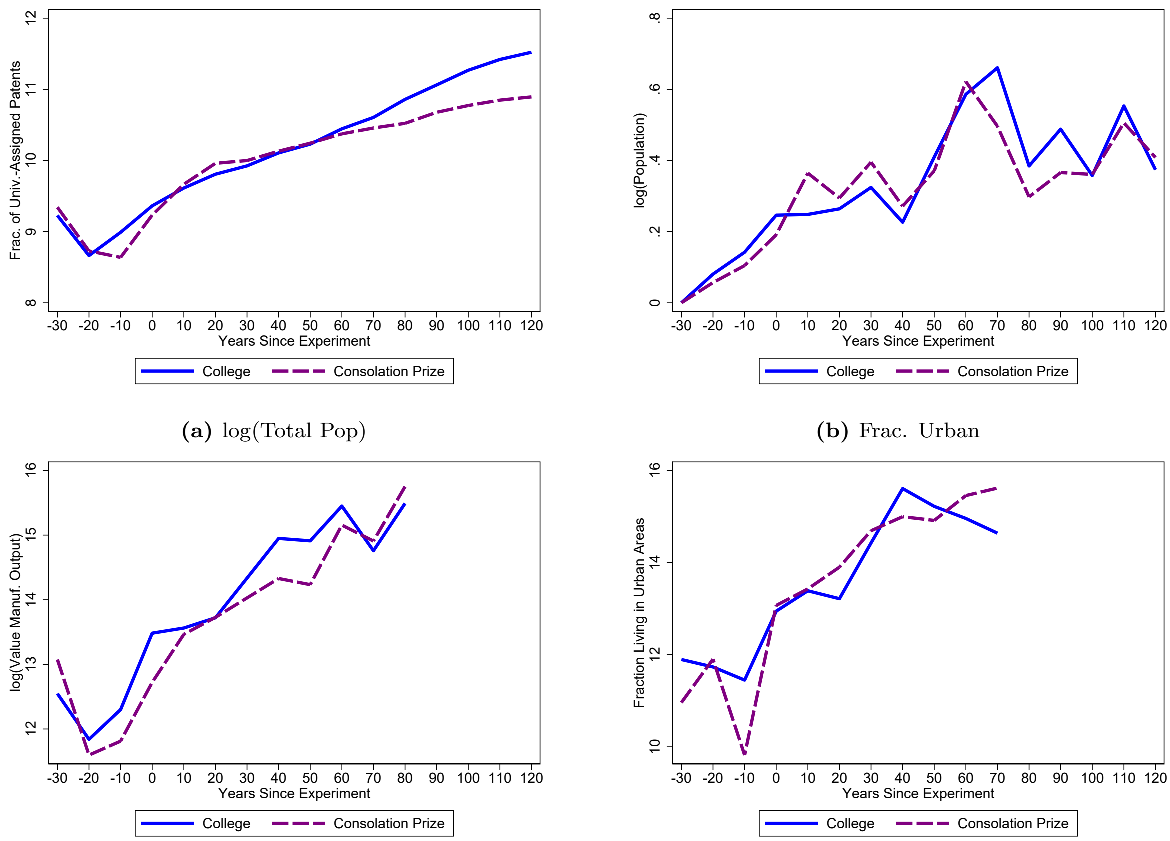
Column 1 lists the college, Column 2 the county and state, Column 3 the type of county (either a college or consolation prize), Column 4 the number of patents over the years 2000-2010, and Column 5 the county population in the 2010 census. Consistent with Figure 7, by 2000-2010 the college counties do tend to be a bit larger and have more patents than the consolation prize counties, although this is not true in all cases and the magnitudes vary substantially. For instance, the consolation prize counties for the University of South Dakota and West Virginia University have more patents over the most recent full decade than do their college counties, while the consolation prize counties for the University of Kansas, University of North Dakota, Oregon State University, University of South Dakota, Utah State University, West Virginia University, and University of Wyoming all have larger populations than their college counties today.

Table A33: Additional Details on Consolation Prize Locations

	College	State	County	Town	Closest City	Distance to City	Drive Time to City
1	Cornell University	New York	Seneca	Ovid	Rochester, NY	63.1 miles	68 minutes
2	University of Kansas	Kansas	Shawnee	Topeka	Kansas City, MO	63.3 miles	59 minutes
3	North Dakota State University	North Dakota	Stutsman	Jamestown	Minneapolis, MN	328 miles	296 minutes
4	Oregon State University	Oregon	Marion	Salem	Portland, OR	46.4 miles	48 minutes
5	University of Colorado	Colorado	Fremont	Canon City	Colorado Springs, CO	44.9 miles	51 minutes
6	University of North Dakota	North Dakota	Burleigh	Bismarck	Minneapolis, MN	427 miles	381 minutes
7	University of New Mexico	New Mexico	San Miguel	Las Vegas	Santa Fe, NM	67.2 miles	66 minutes
8	University of South Dakota	South Dakota	Yankton	Yankton	Omaha, NE	159 miles	147 minutes
9	University of South Dakota	South Dakota	Bon Homme	Bon Homme	Omaha, NE	187 miles	179 minutes
10	University of Wyoming	Wyoming	Uinta	Evanston	Salt Lake City, UT	83.4 miles	82 minutes
11	University of Wyoming	Wyoming	Laramie	Cheyenne	Denver, CO	101 miles	93 minutes
12	Utah State University	Utah	Weber	.	Salt Lake City, UT	42.7 miles	43 minutes
13	West Virginia University	West Virginia	Kanawha	Charleston	Columbus, OH	162 miles	162 minutes

Notes: Details about each of the consolation prize locations. The first column lists the college, the second column the consolation prize's state, the third column the consolation prize's county, the fourth column the consolation prize's town, the fifth column the closest major city to the consolation prize, the sixth column the distance in miles to the major city, and the seventh column the current driving distance from the consolation prize to the major city.

Figure A9: Time Series for Demographic and Economic Variables for Consolation Prizes



(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 consolation prize counties are represented by the dashed line. In each panel, the y-axis is a demographic or economic variable. Data are for the subset of colleges for which a runner-up county received a consolation prize.

Table A34: Consolation Prize Counties: Contemporary Outcomes

	College	County	Type	2000-2010 Patents	2010 Population
1	University of Colorado	Boulder, Colorado	College	4941	290177
2	University of Colorado	Fremont, Colorado	Consolation Prize	7	46941
3	University of Kansas	Douglas, Kansas	College	257	109052
4	University of Kansas	Shawnee, Kansas	Consolation Prize	64	175537
5	New Mexico State University	Donaana, New Mexico	College	127	201670
6	New Mexico State University	San Miguel, New Mexico	Consolation Prize	13	29321
7	Cornell University	Tompkins, New York	College	177	100612
8	Cornell University	Seneca, New York	Consolation Prize	12	35309
9	North Dakota State University	Cass, North Dakota	College	182	144410
10	North Dakota State University	Stutsman, North Dakota	Consolation Prize	50	20984
11	University of North Dakota	Grandforks, North Dakota	College	73	66771
12	University of North Dakota	Burleigh, North Dakota	Consolation Prize	43	78776
13	Oregon State University	Benton, Oregon	College	586	84158
14	Oregon State University	Marion, Oregon	Consolation Prize	366	309894
15	University of South Dakota	Clay, South Dakota	College	12	13816
16	University of South Dakota	Yankton, South Dakota	Consolation Prize	25	22216
17	University of South Dakota	Bon Homme, South Dakota	Consolation Prize	1	7080
18	Utah State University	Cache, Utah	College	531	107078
19	Utah State University	Weber, Utah	Consolation Prize	121	222849
20	West Virginia University	Monongalia, West Virginia	College	52	92715
21	West Virginia University	Kanawha, West Virginia	Consolation Prize	205	192770
22	University of Wyoming	Albany, Wyoming	College	95	34926
23	University of Wyoming	Uinta, Wyoming	Consolation Prize	4	20537
24	University of Wyoming	Laramie, Wyoming	Consolation Prize	41	89221

Notes: Details about the college and consolation prize counties in modern years. The first column lists the college, the second column the county and state, the third column the type of county (either a college or consolation prize), the fourth column the number of patents over the years 2000-2010, and the fifth column the county population in the 2010 census.

G.B.2 Details on the Potential Consolation Prize Cases

In Table A35, I list each of the potential consolation prize experiments. These are cases in which a state established a college with the goal of serving the entire state. While in each of these experiments the state could have also decided to establish a consolation prize institution at the same time in a different location, it did not do so (to the best of my knowledge based on the narrative historical record) in any of these experiments.

Figure A10 replicates Figure A9 but uses the potential consolation prizes and their runner-up counties instead of the actual consolation prize counties. The potential consolation prize college counties were also evolving similarly to their runner-up counties in the years prior to establishing a college. I once again omit confidence intervals for readability; the potential consolation prize college counties are not statistically different from their runner-up counties in the years prior to establishing the college.

G.B.3 Additional Consolation Prize Results

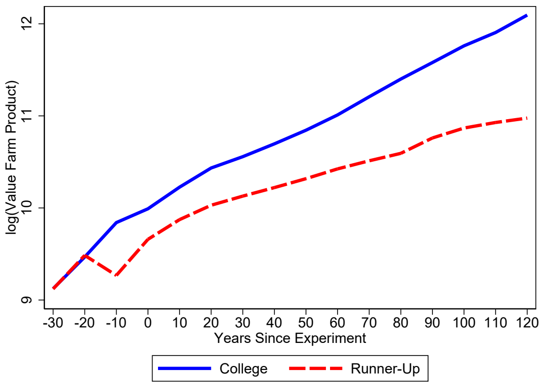
In this section, I present additional results related to the consolation prize cases. First, I plot patenting and population in the college counties relative to the consolation prizes, broken up by the three types of consolation prizes. Figure A11 replicates Panel (a) of Figure 7, except each panel in Figure A11 shows results for a different type of consolation prize. Panel (a) plots patenting in the college and consolation prize counties for the cases in which the consolation prize is a state capital, Panel (b) plots results for the cases in which the consolation prize is an asylum, and Panel (c) plots results for the cases in which the consolation prize is a penitentiary. Figure A12 is similar except that it plots population rather than patenting,

Table A35: List of Potential Consolation Prizes

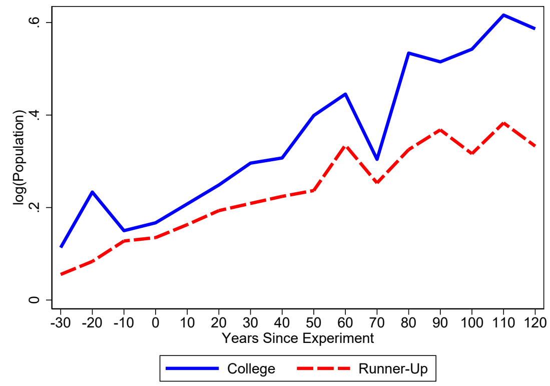
	College	State	College County
1	Auburn University	Alabama	Lee
2	Arkansas Tech University	Arkansas	Sebastian
3	University of Arkansas	Arkansas	Washington
4	University of California Berkeley	California	Contracost
5	University of California Davis	California	Solano
6	University of Florida	Florida	Columbia
7	Georgia Institute of Technology	Georgia	Greene
8	University of Idaho	Idaho	Bonneville
9	University of Illinois	Illinois	Champaign
10	Purdue University	Indiana	Tippecanoe
11	Iowa State University	Iowa	Marshall
12	Louisiana State University	Louisiana	Eastbatonr
13	University of Maine	Maine	Sagadahoc
14	University of Mississippi	Mississippi	Winston
15	Missouri University of Science and Technology	Missouri	Iron
16	University of Missouri	Missouri	Cole
17	University of Nevada	Nevada	Carsoncity
18	University of New Hampshire	New Hampshire	Strafford
19	East Carolina University	North Carolina	Edgecombe
20	North Carolina State University	North Carolina	Wake
21	University of Oregon	Oregon	Lane
22	Pennsylvania State University	Pennsylvania	Centre
23	Clemson University	South Carolina	Richland
24	University of Tennessee	Tennessee	Knox
25	Texas A and M University	Texas	Austin
26	Texas Tech	Texas	Nolan
27	University of Texas Austin	Texas	Travis
28	Virginia Polytechnic Institute	Virginia	Rockbridge
29	Washington State University	Washington	Whitman
30	University of Wisconsin	Wisconsin	Dane

Notes: The "potential consolation prize colleges" consisting of any land grant, technical, or other public college that was established and designed to serve the needs of the entire state but in which no runner-up county received a consolation prize. The first column lists the college, the second column the college's state, and the third column the college's county.

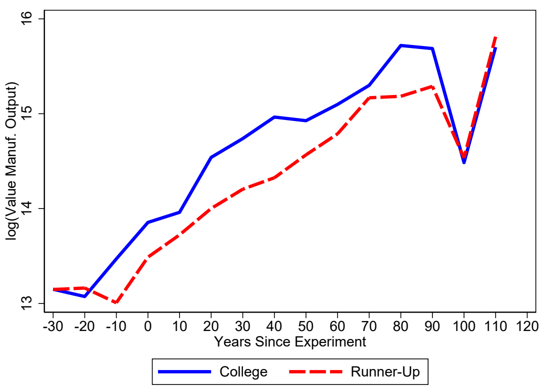
Figure A10: Time Series for Demographic and Economic Variables for Potential Consolation Prizes



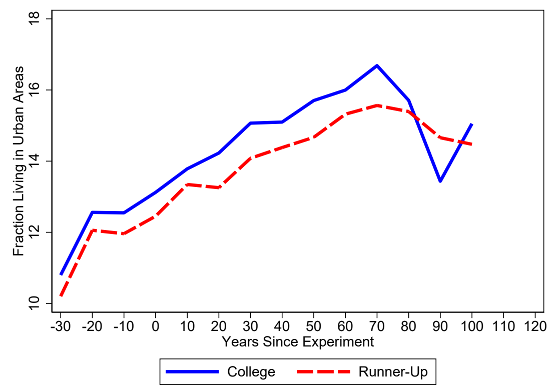
(a) log(Total Pop)



(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. In each panel, the y-axis is a demographic or economic variable. Data are for the "potential consolation prize colleges" consisting of any land grant, technical, or other public college was established and designed to serve the needs of the entire state but in which no runner-up county received a consolation prize.

replicating Panel (b) of Figure 7.

Figure A13 plots dynamic difference-in-difference results for all consolation prizes, for the potential consolation prizes, and for the three types of consolation prizes separately, replicating Figure 4 with the consolation prize sample.[30] Consistent with the results in Table 8 and Figures 7 and A11, panel (a) shows that when considering all consolation prize experiments, the college counties are only significantly different from the consolation prize counties in the most recent decade, 12 decades after the college was established; in most decades the difference is small in magnitude or even negative. When considering the potential consolation prize experiments instead in panel (b), colleges have statistically significantly more patents than the potential consolation prizes after seven decades have passed, and the estimates continue to be individually statistically significant for every decade thereafter. Similar to panel (a), magnitudes are also typically small—and standard errors understandably even larger—when considering the state capital, asylums, or prison consolation prizes separately in panels (c)-(e), respectively.

I next show how establishing a college increases population in the college counties relative to the consolation prize counties, replicating the results in Column 1 of Table 4 using the samples of colleges from each column of Table 8. Results are presented in Table A36. As expected, while the college counties have a large increase in logged population relative to the potential consolation pries (Column 1), college counties show a smaller and statistically insignificant increase in logged population relative to the consolation prizes (Column 2). The coefficient when comparing colleges to state capitals is close to zero (Column 3) and is larger

[30]Similar dynamic results where population or patenting per capita are the dependent variables, analogous to Figure 5, are available upon request.

when comparing colleges to asylums (Column 4) or penitentiaries (Column 5), although all three are smaller than the coefficient in Column 1 and are statistically insignificant.

Figures 7 and A11-A13 confirm the conclusion in Section IV.B that establishing a college increases local patenting at best only modestly compared to counties that receive a consolation prize. In spite of this, Table A34 reveals that, in many cases, college counties have many more patents over the period 2000-2010 than do their respective consolation prizes. As suggested by both Figure 7 and panel (a) of Figure A13, the difference between college and consolation prizes widened substantially just in the past decade. Is it misleading to claim that colleges have a modest effect relative to consolation prizes given the large observed contemporary differences? To explore this issue more explicitly, in Columns 1 and 2 of Table A37 I estimate what is essentially a long-differences specification. I estimate the baseline specification using only observations from the decade prior to the establishment of each college and from 2000-2010. In Column 1, I include only the consolation prize experiments. The estimated coefficient is 143.5 log points, or about 320% more patents in college counties today relative to their consolation prizes. At face value, this magnitude is large, and in fact is larger than the baseline estimate from all experiments in Table 2 and more than ten times the magnitude of the baseline difference-in-difference result for the consolation prize sample in Column 2 of Table 8. Columns 2 of Table A37 puts this magnitude into perspective, however. When examining the potential consolation prize sample in Column 2, the estimated coefficient is a strongly statistically significant 189.7 log points, or about 566.5%, which is 1.8 times as large as the percentage increase for the consolation prize sample. Hence, for both subsamples of colleges the estimated magnitude is much larger in the long differences specification than it is in the baseline difference-in-difference (consistent with colleges playing a larger role in innovation today as suggested in Section III.A and F.C), but even today the magnitude for the consolation prize colleges is substantially less than that of comparable non-consolation prize colleges. Moreover, this effect becomes noticeable for the consolation prize colleges only in very recent decades, more than a century after the colleges were initially established and much later than the non-consolation prize samples. To further underscore this point, Columns 3 and 4 of Table A37 repeat this long-difference specification but use patents from the years 1980-1990 instead of 2000-2010. In this specification, which includes the first full decade post-Bayh-Dole Act, the estimated coefficient is a statistically insignificant 0.696 log points, or about 101%, while it is a highly statistically significant 134.5 log points, about 284%, for the potential consolation prizes. Finally, while the baseline sample and potential consolation prize samples show nearly monotonic increases in patenting in the college county relative to their runners-up beginning after about four decades, panel (b) of Figures 7 and A13 show that the difference between college and consolation prizes grows and shrinks several times over the last 120 years, raising the possibility that the sizable difference in patenting between college and consolation prizes today is transitory.

Next, I show that the consolation prize results in Table 8 are robust to several alternative specifications. First, I omit the three cases described in Section G.B.1 above in which the consolation prize was located reasonably close to a major city. These occurred for Utah State, University of Colorado, and Oregon State; in the first a consolation prize was located about 40 miles from Salt Lake City, in the second the consolation prize was located 45 miles from Colorado Springs, and in the third a consolation prize was located about 46 miles from Portland. I present results omitting these three experiments from the analysis in Column 1 of Table A38. When omitting these cases, the difference between the college and consolation prize counties is even smaller than in the baseline estimates, at only 17.3 log points.

Next, I build on the intuition from Figures 7 and A13 to show that the college and consolation prize counties were particularly similar to one another for the first several decades after establishing the college. To do this, I discard all data from 1980 or later; as noted above, 1980 saw the passage of the Bayh-Dole Act, which accelerated trends toward encouraging colleges to patent faculty inventions. Consistent with the discussion above, without the most recent three decades of data, the difference between the college and consolation prize counties is close to zero in magnitude: college counties see only 2.3 log points (2.3%) more patents per year than the consolation prize counties.

In Column 3 I exclude all college consolation prize counties for experiments in years after the consolation prize county received a college of its own. Consolation prize counties may be especially likely to later receive a college of their own because not only were they considered suitable sites for a college initially, but the presence of the consolation prize drove population growth which may have driven up local demand for a college. It is therefore necessary to ensure that the similarity between the college and consolation prize counties are not driven by the establishment of a college in the consolation prize counties. This exercise is similar in spirit to that from Section E.A, although in this case the sample of consolation prize counties is small enough that I can manually verify when each college opened in the consolation prize counties. For instance, in Stutsman County, ND, the University of Jamestown was established in 1883, so I drop all counties in the North Dakota State University experiment starting in 1883. Burleigh County, ND, saw the establishment of Bismarck State College in 1939, so I drop all counties in the University of North Dakota experiment after 1939. The full list of dates the first college was established in each consolation prize county is available upon request. When excluding these counties in these years, the coefficient is similar to the baseline estimate.

In Column 4, I re-estimate the baseline consolation prize specification (Column 2 from Table 8) while controlling for logged population. In this case, the difference between the college and consolation prize counties is still statistically insignificant and smaller in magnitude than the baseline estimate. In Columns 5-7, I repeat Columns 3-5 from Table 8 in which I consider cases in which the consolation prizes are state capitals, asylums, and penitentiaries, respectively, while controlling for county population. For state capitals and penitentiaries, the coefficients are smaller than the corresponding estimates in Table 8) and still statistically insignificant. For asylums, however, controlling for population results in a larger estimate that is statistically significant at the 5% level. For all four of these columns, I stress that population is an endogenous outcome variable, and so the results must be interpreted with caution.

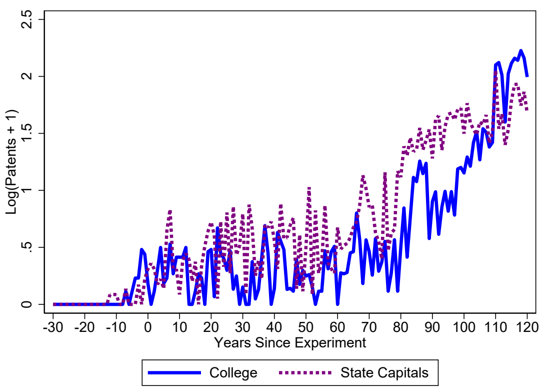
Table A36: Consolation Prizes and Population

	Potential Consolation Prizes	Consolation Prizes	State Capitals	Asylums	Penitentiaries
College * PostCollege	0.633***	0.352	0.023	0.498	0.485
	(0.190)	(0.242)	(0.263)	(0.320)	(0.466)
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Year Range	1840-2010	1840-2010	1840-2010	1840-2010	1840-2010
County-Year Observations	1,620	432	270	198	162
# Counties	87	24	15	11	9
# Experiments	30	11	6	4	4
Adjusted R-Squared	0.647	0.625	0.585	0.613	0.554

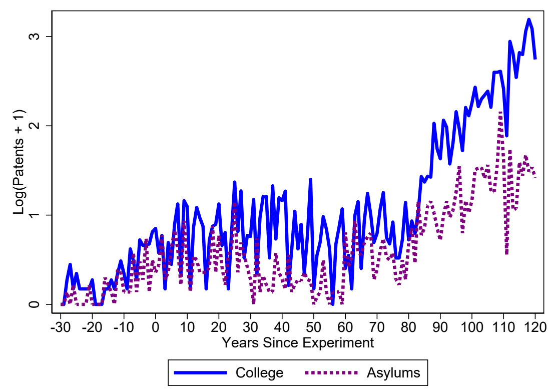
Notes: Column 1 estimates the effect of establishing a college on local population for the sample of potential consolation prize experiments. Column 2 estimates the effect of establishing a college on local population for the sample of actual consolation prize experiments. Columns 3-5 repeat Column 2 but use only the college experiments in which the consolation prize is a state capital, asylum, or penitentiary, respectively. The dependent variable in all columns is log($Population$). Results are for high quality experiments only. Standard errors are clustered by county and shown in parentheses. Stars indicate statistical significance: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Figure A11: Patenting in College and Consolation Prize Counties by Type of Consolation Prize

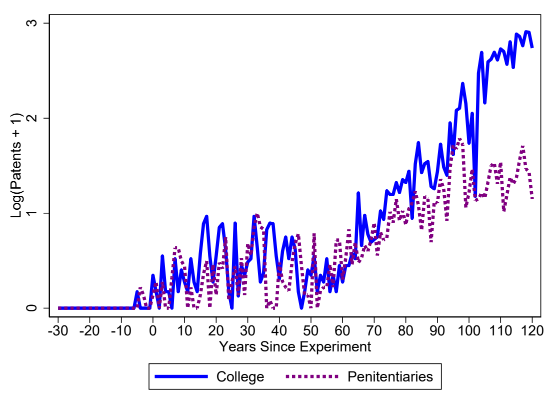
(a) State Capitals



(b) Asylums



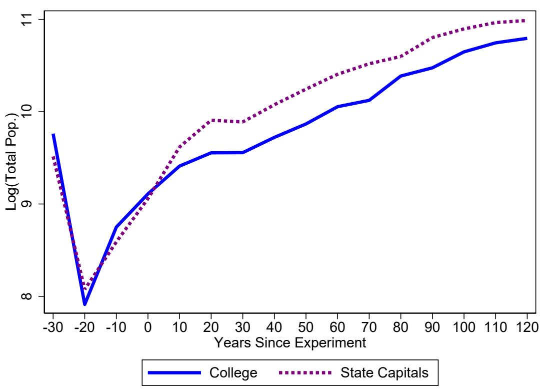
(c) Penitentiaries



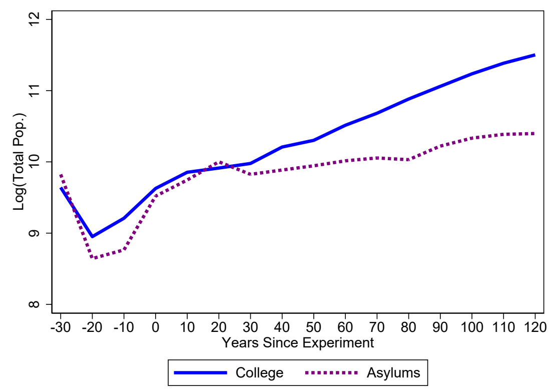
Notes: The x-axis shows the number of years since the college experiment. The year of the establishment of the new college is normalized to 0. Everything left of 0 shows pre-college results; everything to the right shows post-college results. Panel (a) uses the sample in which the consolation prize is a state capital, (b) the sample in which the consolation prize is an asylum, (c) the sample in which the consolation prize is a penitentiary. In all panels, the y-axis is log(Patents + 1). The college counties are represented by the solid line, while the consolation prize counties are represented by the dashed line.

Figure A12: Population in College and Consolation Prize Counties by Type of Consolation Prize

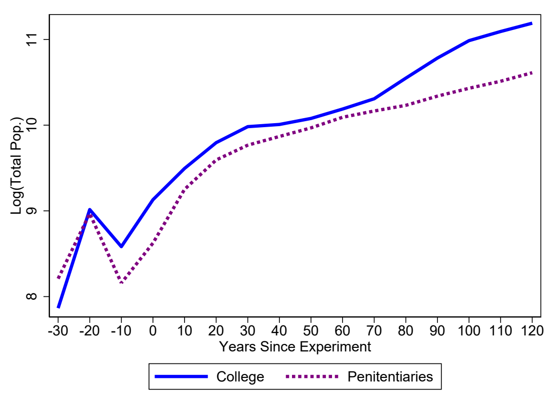
(a) State Capitals



(b) Asylums

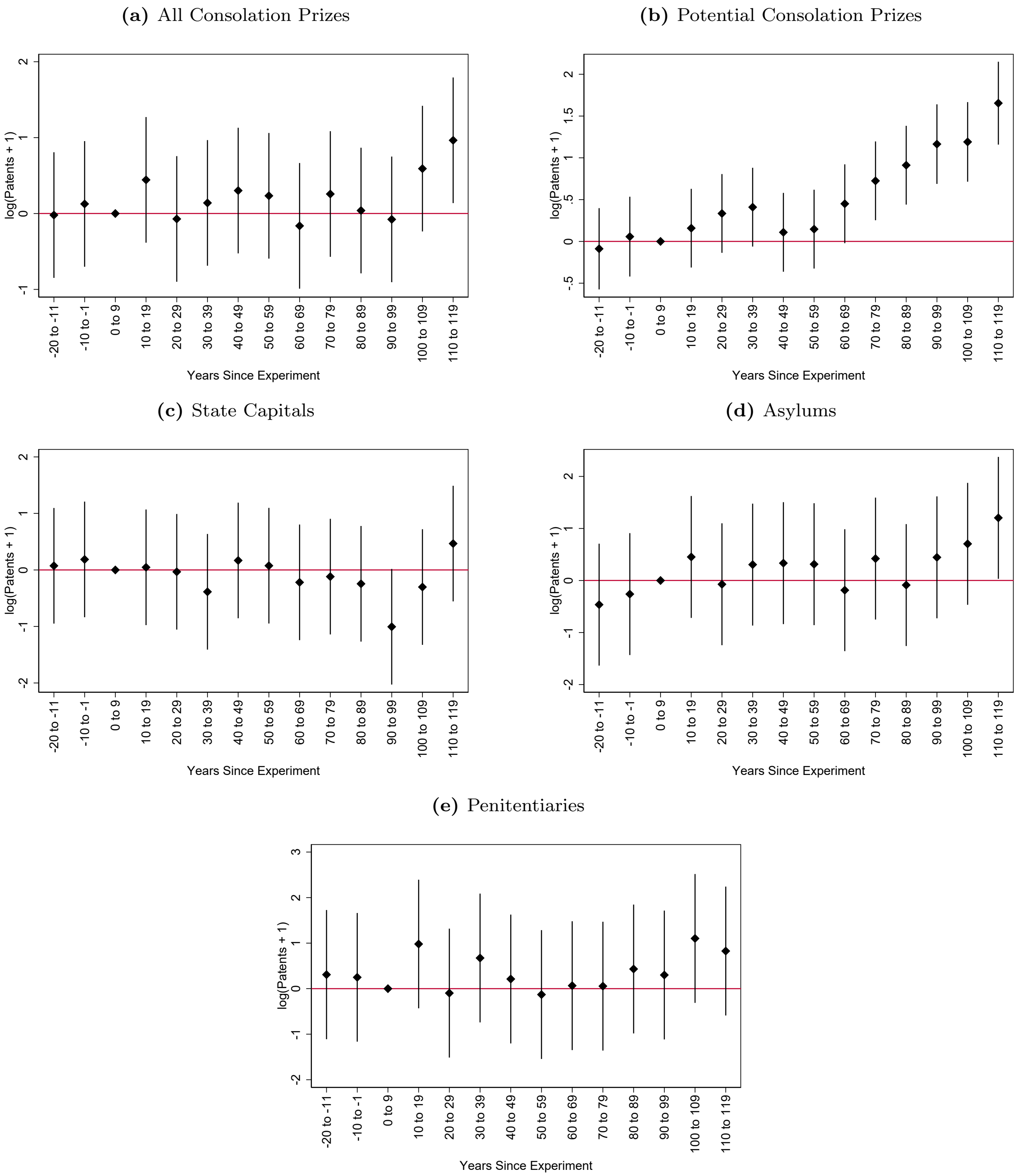


(c) Penitentiaries



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. Panel (a) uses the sample in which the consolation prize is a state capital, (b) the sample in which the consolation prize is an asylum, (c) the sample in which the consolation prize is a penitentiary. In all panels, the y-axis is log(Population). The college counties are represented by the solid line, while the consolation prize counties are represented by the dashed line.

Figure A13: Dynamics of the Effect of Local Colleges for the Consolation Prize Samples



Notes: Estimated coefficients of the shift in logged patenting in college counties 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. The black diamonds show coefficients comparing the college counties to runner-up counties. The gray triangles show coefficients comparing the college counties to the non-experimental counties. Data are for high quality experiments only. Panel (a) uses the sample of all consolation prize experiments, (b) the sample of all potential consolation prize experiments, (c) the sample in which the consolation prize is a state capital, (d) the sample in which the consolation prize is an asylum, and (e) the sample in which the consolation prize is a penitentiary.

Table A37: Long Difference Results

	2000-2010		1980-1990	
	Consolation Prizes	Potential Consolation Prizes	Consolation Prizes	Potential Consolation Prizes
College * PostCollege	1.435**	1.897***	0.696	1.345***
	(0.521)	(0.371)	(0.458)	(0.308)
Experiment FE	Yes	Yes	Yes	Yes
Year Range	1852-2010	1836-2010	1852-1990	1836-1990
County-Year Observations	504	1,813	504	1,813
# Counties	24	87	24	87
# Experiments	11	30	11	30
Adjusted R-Squared	0.795	0.800	0.727	0.757

Notes: "Long difference" results in which data are from the ten years before each college is established and the years 2000-2010. Column 1 estimates the long difference effect of establishing a college on local patenting for the sample of actual consolation prize experiments. Column 2 estimates the long difference effect of establishing a college on local patenting for the sample of potential consolation prize experiments. Column 3 estimates the long difference effect of establishing a college on local patenting for the baseline sample of colleges. The dependent variable in all columns is $\log(Patents + 1)$. 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 A38: Additional Consolation Prize Results

	No Consolation Prizes Near Large Cities	No Post-1980 Years	No Consolation Prizes that Get a College	Control for Population			
				All Consolation Prizes	State Capitals	Asylums	Penitentiaries
College * PostCollege	0.173	0.023	0.272	0.143	-0.200	0.464**	0.289
	(0.219)	(0.185)	(0.344)	(0.191)	(0.144)	(0.194)	(0.485)
log(Population)				0.665***	0.706***	0.304***	1.207***
				(0.141)	(0.134)	(0.070)	(0.351)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-1979	1836-2010	1840-2010	1840-2010	1840-2010	1850-2010
County-Year Observations	3,150	3,456	2,431	378	218	142	133
# Counties	18	24	24	24	14	9	9
# Experiments	8	11	11	11	6	4	4
Adjusted R-Squared	0.602	0.524	0.563	0.675	0.653	0.652	0.695

Notes: Additional results comparing college counties to runner-up counties that receive a consolation prize. The dependent variable in all columns is $log(Patents + 1)$. In Column 1, I omit experiments in which the consolation prize county is established near a large city. In Column 2, I omit all years 1980 and after. In Column 3, I exclude experiments in years after which a consolation prize county received a college of its own. In Column 4, I compare college to consolation prize counties while controlling for logged county population. In Columns 5-7, I present results only for the subsample of experiments in which the consolation prize is a state capital, an asylum, or a penitentiary, respectively, 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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

G.C Heterogeneity by College Type

I further break down the results of different types of colleges on patenting. In Column 1 of Table A39, I control for logged county population, since different types of colleges may drive migration in different ways. After controlling for population, neither coefficient is statistically significant; the coefficient for practical colleges is close to zero in magnitude, while that for classical colleges is -17.7 log points (-19.4%). 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 benefit of this classification is that it uses all of the colleges in the sample. But because colleges like normal schools and HBCUs may differ from the land grant colleges and other public institutions in other dimensions beyond just their curricula (for instance, they may have fewer resources or political support), the comparison between the two estimates is more difficult to interpret. In this specification, the estimate for practical colleges is larger than that in Table 8 and still statistically significant, while the estimate for classical colleges is a bit smaller and still statistically insignificant. The difference between practical and classical colleges is still qualitatively the same when the alternative definitions are used, however, and the two coefficients are not statistically different from one another.

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, the largest coefficient is for military academies; while these school focus on technical training, their graduates enter military service and leave their local counties after graduating. Other private colleges have the next largest coefficient, even though this group comprises small liberal arts colleges. Technical schools and and land grant colleges, which focus on technical skills, are in the middle of the pack. 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 positive and statistically significant 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 be larger and hence cause more population growth. Indeed, after controlling for population, the effect of both types of colleges are smaller 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 coefficient estimates do not decline monotonically with quality: the coefficient is largest for schools ranked 1-75, then those ranked 76-150, but the unranked schools have a larger coefficients than those ranked 151-225; only the first two groups are individually statistically significant and most are not statistically different from one another. 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 6 I control for logged county population. The coefficients one again do not 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.

G.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.

Table A39: Additional Results by College Type

	Control for Population	Alternative Practical vs. Classical	Public vs. Private	Public vs. Private	College Rank	College Rank
Practical College Interaction	0.033	0.600***				
	(0.168)	(0.154)				
Classical College Interaction	-0.177	0.339				
	(0.216)	(0.266)				
log(Total Population)	0.870***			0.864***		0.865***
	(0.067)			(0.066)		(0.063)
All Public Colleges			0.451***	0.027		
			(0.153)	(0.109)		
All Private Colleges			1.289	0.514		
			(0.860)	(0.321)		
Rank 1-75					0.971***	0.478***
					(0.277)	(0.115)
Rank 76-150					0.600***	0.107
					(0.219)	(0.331)
Rank 151-225					0.069	-0.251
					(0.218)	(0.239)
Unranked					0.459	-0.037
					(0.310)	(0.138)
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
County-Year Observations	30,919	33,425	33,425	30,919	33,425	30,919
# Counties	174	174	174	174	174	174
# Experiments	63	63	63	63	63	63
Adjusted R-Squared	0.730	0.611	0.611	0.730	0.614	0.731

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: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

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 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

$$\begin{aligned} PatentMeasure_{ijt} =& \delta_1 College_i * PostCollege_{it} * Above75thPct_i \\ &+ \delta_2 College_i * PostCollege_{it} * Below25thPct_i \\ &+ County_i + Year_t + \epsilon_{ijt}, \end{aligned} \tag{9}$$

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 A40 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 a statistically significant 118.5 log points more patents per year in the top quartile of college counties relative to their runner-ups, compared to 32 log points more patents per year in the bottom quartile of college counties relative to their runner-ups (statistically significant at the 10% level). Column 2 shows heterogeneity by the fraction of the county population living in an urban area. Once again, more urbanized counties have a large and statistically significant coefficient, while the coefficient for the least urbanized counties is smaller in magnitude but still statistically significant. This pattern is repeated in Column 3, when I split effects by prior county patenting: counties in the top quartile of patenting see a 111.5 log point increase in local patenting relative to their runner-up counties, while counties in the bottom quartile of patenting see a 34.4 log point increase in patenting. 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 (176.2 log points, significant at the 1% level) than do the top quartile of counties (61.9 log points, also statistically significant at the 1% level), although data on railroad access is not available for all 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 more developed. Taking the above estimates at face value, a college has the largest effect when a county is large, urbanized, and inventive, but lacks transportation infrastructure. This conclusion is also consistent with the results in Section E.C, which finds that the effect of a college is larger when there is a preexisting college in the county, although the difference is modest. 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.

Table A40: Heterogeneity by County Characteristics

	By Previous			
	County Population	Fraction Urbanized	County Patents	Access to Railroads
Above 75th Pct.	1.185***	0.827***	1.115***	0.619***
	(0.245)	(0.290)	(0.272)	(0.032)
Below 25th Pct.	0.320*	0.452***	0.344***	1.762***
	(0.180)	(0.091)	(0.087)	(0.032)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Year Range	1836-2010	1836-2010	1836-2010	1836-2010
County-Year Observations	472,516	713,882	757,691	30,114
# Counties	1,139	1,379	1,605	195
# Experiments	30	43	51	2
Adjusted R-Squared	0.471	0.427	0.450	0.519

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, Daron, Ufuk Akcigit, and Murat Alp Celik. 2014. "Young, restless and creative: openness to disruption and creative innovation." NBER Working Paper.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017. "The rise of American ingenuity: innovation and inventors of the golden age." Unpublished, University of Chicago.

Andrews, Michael J. 2021*a*. "Historical appendix: How do institutions of higher education affect local invention? Evidence from the establishment of U.S. colleges." Unpublished, UMBC.

Andrews, Michael J. 2021*b*. "Historical patent data: a practitioner's guide." *Journal of Economics and Management Strategy*, 30(2): 368–397.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion.* Princeton:Princeton University Press.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2020. "How well do automated linking methods perform in historical samples? Lessons from US historical data." *Journal of Economic Literature*, 58(4): 997–1044.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who becomes an inventor in America? The importance of exposure to innovation." *Quarterly Journal of Economics*, 134(2): 647–713.

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

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

Blundell, Richard, Rachel Griffith, and John Van Reenen. 1995. “Dynamic count data models of technological innovation.” *The Economic Journal*, 105(429): 333–344.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. “Mobility report cards: the role of colleges in intergenerational mobility.” Unpublished, NBER Working Paper No. 23618.

Fagan, George V. 2006. *Air Force Academy heritage: the early years.* Golden, CO:Fulcrum Publishing.

Ferrie, Joseph P. 1996. “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, Claudia, and Lawrence F. Katz. 1999. “The shaping of higher education: the formative years in the United States, 1890 to 1940.” *Journal of Economic Perspectives*, 13(1): 37–62.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. “The NBER patent citation data file: lessons, insights and methodological tools.” NBER Working Paper.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg. 2005. "Market value and patent citations." *RAND Journal of Economics*, 36(1): 16–38.

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

Jung, Taehyun, and Olof Ejermo. 2014. "Demographic patterns and trends in patenting: gender, age, and education of inventors." *Technological Forecasting and Social Change*, 86: 110–124.

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

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

Kuhn, Jeffrey M., and Neil 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, Jeffrey M., Kenneth A. Younge, and Alan 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.* Vol. 1-3, Washington, DC:Government Printing Office.

Logan, Trevon D., and John M. Parman. 2017. "The national rise in residential segregation." *Journal of Economic History*, 77(1): 127–170.

Maloney, William F., and Felipe Valencia Caicedo. 2017. "Engineering growth: innovative capacity and development in the Americas." Unpublished, IZA.

Marco, Alan C., Michael Carley, Steven Jackson, and Amanda F. Myers. 2015. "The USPTO historical patent data files: two centuries of invention." Unpublished, USPTO Economic Working Paper No. 2015-1.

Miller, Jim. 2016*a*. "Jim Shaw Historical Patent Data." electronic correspondence, October 24, 2016.

Miller, Jim. 2016*b*. "Jim Shaw Historical Patent Data." electronic correspondence, November 1, 2016.

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

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

Nauman, Robert Allen. 2004. *On the wings of modernism: the United States Air Force Academy.* Urbana and Chicago, IL:University of Illinois Press.

Petralia, Sergio, Pierre-Alexandre Balland, and David L. Rigby. 2016*a*. "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, Sergio, Pierre-Alexandre Balland, and David L. Rigby. 2016*b*. "Historical Patent Dataset." `http://dx.doi.org/10.7910/DVN/BPC15W`, accessed September 15, 2016.

Royston, Patrick, and Douglas 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, Bhaven N. 2006. "Patenting and US academic research in the 20th century: the world before and after Bayh-Dole." *Research Policy*, 35(6): 772–789.

Sarada, Michael J. Andrews, and Nicolas L. Ziebarth. 2019. "Changes in the demographics of American inventors, 1870-1940." *Explorations in Economic History*, 74.

Trajtenberg, Manuel. 1990. "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)." Office for Patent and Trademark Information, U.S. Patent and Trademark Office Technology Assessment and Forecasting Report, 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.

Vara, Vauhini. 2015. "How Utah became the next Silicon Valley." *The New Yorker*.

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

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

Zolas, Nikolas, Nathan Goldschlag, Ron Jarmin, Paula Stephan, Jason Owen-Smith, Rebecca F. Rosen, Barbara McFadden Allen, Bruce A. Weinberg, and Julia I. Lane. 2015. "Wrapping it up in a person: examining employment and earnings outcomes for Ph.D. recipients." *Science*, 350(6266): 1367–1371.