University Spillovers: Productivity in the Agricultural and Manufacturing Sectors 1870-1940

David Courtright

Clemson University, David.Court17@yahoo.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation
https://tigerprints.clemson.edu/all_theses/3277

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
UNIVERSITY SPILOVERS: PRODUCTIVITY IN THE AGRICULTURAL AND MANUFACTURING SECTORS 1870-1940

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
Economics

by
David Courtright
May 2020

Accepted by:
Dr. Jorge Garcia, Committee Chair
Dr. Robert Fleck
Dr. Scott Templeton
Abstract

This paper estimates the effects of college attendance from 1870 to 1940 on productivity in the manufacturing and agricultural sectors. I use data from the Records of the Office of Education, Census of Manufacturers, and Census of Agriculture from 1870 to 1940 to analyze trends in college attendance and productivity in the manufacturing and agricultural sectors. I find that every one student increase in the average number of students enrolled in a university in a county over the previous decade increases agricultural output per acre by $.01 and manufacturing output per employee by $2.13.
Acknowledgments

I am deeply indebted to Jorge Garcia for many helpful conversations and insightful advice. I would like to thank Jorge Garcia, Robert Fleck, and Scott Templeton for serving on my thesis committee. I would also like to thank Jorge Garcia, Babur De los Santos, Jeremy Atack, the Minneapolis Federal Reserve Bank, Steven Manson, Jonathan Schroeder, David Van Riper, Steven Ruggles, Fabian Eckert, Andrés Gvirtz, Jack Liang, and Michael Peters for sharing data and/or code. In addition to Jorge Garcia, this paper benefited from conversations with Babur De los Santos, Robert Fleck, Ian Davis, and Shawn Kantor. Finally, I would like to thank Jorge Garcia, Babur De los Santos, Robert Fleck, Scott Templeton, and Reed Watson for advice throughout my time at Clemson University.
# Contents

Title Page ................................................................. i  
Abstract ................................................................. ii  
Acknowledgments ............................................................ iii  
List of Tables .............................................................. vi  
List of Figures .............................................................. vii  
1 Introduction ............................................................... 1  
2 Expansion of the University System ................................. 3  
3 Data and Descriptive Statistics ........................................ 8  
   3.1 Data Sources .......................................................... 8  
   3.2 Sample Selection and Summary Statistics ......................... 11  
4 Empirical Model .......................................................... 18  
   4.1 Defining Productivity ............................................... 18  
   4.2 Estimating Equation ............................................... 19  
5 Empirical Results ........................................................ 20  
   5.1 Estimated Impacts on Productivity ................................ 20  
   5.2 Theoretical Explanation ........................................... 23  
   5.3 Measurement Error ............................................... 23  
   5.4 Endogeneity Problem ............................................... 26  
   5.5 Robustness Checks ................................................ 27  
6 Conclusion ................................................................. 31  
   6.1 Limitations Of Study ................................................. 31  
   6.2 Recommendations for Further Research ............................ 32  
Bibliography ............................................................... 32
<table>
<thead>
<tr>
<th>Appendices</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Location Maps</td>
<td>36</td>
</tr>
<tr>
<td>B Maps of Agricultural Statistics</td>
<td>69</td>
</tr>
<tr>
<td>C Map of Manufacturing Statistics</td>
<td>89</td>
</tr>
<tr>
<td>D Additional Graphs</td>
<td>110</td>
</tr>
</tbody>
</table>
List of Tables

1  Summary Statistics of College Variables ........................................ 12
2  Summary Statistics of Firm Variables ........................................... 14
3  Fixed Effect Regression Table for Productivity Lagged Students ...... 21
# List of Figures

1. Location of Colleges ........................................ 5
2. Geographic Distribution of Colleges .......................... 6
3. Geographic Distribution of Students .......................... 7
4. Geographic Distribution of Agricultural Output Per Acre (USD) ... 16
5. Geographic Distribution of Manufacturing Output Per Employee (USD) 17
6. Actual and Counterfactual Productivity Measures ............... 22
7. Location of Colleges in the U.S. ............................. 37
8. Geographic Distribution of Railroads ......................... 41
9. Geographic Distribution of Navigable Rivers .................. 44
10. Geographic Distribution of Canals ............................ 47
11. College Locations with Access to Transportation Networks .... 50
12. Geographic Distribution of Colleges .......................... 53
13. Geographic Distribution of College Openings ................ 57
14. Geographic Distribution of Students .......................... 61
15. Geographic Distribution of Professors ........................ 65
16. Geographic Distribution of Farms ............................ 70
17. Geographic Distribution of Wages Paid to Farmers (USD) ...... 74
18. Geographic Distribution of Agricultural Revenue (USD) ........ 77
19. Geographic Distribution of Farm Land (Acres) ................ 81
20. Geographic Distribution of Agricultural Revenue Per Acre (USD) . 85
21. Geographic Distribution of Manufacturing Establishments ...... 90
22. Geographic Distribution of Wages Paid to Manufacturing Employees (USD) ........................................ 94
23. Geographic Distribution of Revenue From Manufacturing Activity (USD) 98
24. Geographic Distribution of Manufacturing Employees ........ 102
25. Geographic Distribution of Manufacturing Revenue Per Employees (USD) 106
26. Outcome Variable and Lagged Number of Students ............ 111
Chapter 1

Introduction

The period from the mid-19th century to the mid-20th century was a time of massive growth in the university system of the United States (Goldin and Katz, 1999). Universities increased in number, size, and complexity. These institutions impact the local economy through a variety of direct and indirect mechanisms. This period was also a time of dramatic changes in the American economy. Manufacturing and agriculture grew substantially and the US established itself as the world’s industrial leader.

Estimating the effects of universities on productivity is made difficult by two central challenges. First, unobserved factors may affect both universities and firms leading to the colocation of colleges and highly productive firms. Second, the effects of universities on productivity may take many decades to manifest.

In many cases, the historical record shows that the founding of a college may be plausibly exogenous (Moretti, 2004). For example, Georgia sought to establish a technical college in 1886 and was deciding whether it should be located in Macon or Atlanta. Both were similar on a number of observable characteristics including geographic location and access to transportation networks. The site selection decision
required 24 ballots for Atlanta to acquire a majority and, therefore, it is plausible Georgia Tech could have been located in Macon (Andrews, 2019). The inclusion of fixed effects in my model additionally controls for county and county-by-year unobserved heterogeneity.

In this paper, I examine how the growth in university attendance, measured as average enrollment in universities in a county over the current and previous 9 years, impacted the relative productivity of agricultural and manufacturing firms, measured as output per acre and output per employee, respectively. My analysis employs data sources on manufacturing, agriculture, colleges, and population that I link through location data.

In reduced-form regressions, I find that every one student increase in the average number of students enrolled in a university in a county over the previous decade increases agricultural output per acre by $0.01 and manufacturing output per employee by $2.13. Additionally, I find that the effect of universities on agricultural productivity occurs as early as 1880 but the effect on manufacturing productivity takes longer to manifest and does not appear significant until 1920.

This paper contributes to a variety of different literatures associated with human capital transmission and regional development. First, I use a comprehensive data set on US colleges that allows me to better estimate the effect of universities by exploiting variation in the size of student populations where much of the previous literature focuses on the existence of a college as the explanatory variable. Second, I examine the effects on both manufacturing and agricultural productivity whereas most of the previous literature focuses on one or the other.

This thesis unfolds as follows. Section II discusses the expansion of colleges in the US, Section III describes my data, Section IV describes my empirical model, Section V describes my findings, Section VI concludes.
Chapter 2

Expansion of the University System

The period from the mid-19th century to the mid-20th century saw the development of many of the defining features of higher education today— the large average size of institutions, the coexistence of liberal arts colleges and research universities, and the substantial share of enrollment in the public sector (Goldin and Katz, 1999). I find that from 1870 to 1940, the number of colleges in the Reports of the Commissioner of Education more than doubled from 332 to 700 with a peak of 906 colleges in 1926. In the same period mean student enrollment in a county with a college increased nearly tenfold from 244.7 to 2,387.96.

This period saw increased governmental support of the university system. The Morrill Act of 1862 granted each state 30,000 acres for each of its congressional seats to be sold in order to finance the establishment of colleges specializing in agriculture and the mechanic arts (Cunningham, 2019). A similar bill in 1890 provided regular funds to support land-grant colleges. Altogether, 73 land-grant colleges and universities were founded, with each state having at least one (Moretti, 2004). These state
institutions increasingly focused on providing public goods for the state in the form of knowledge. This had two major implications. First, state institutions in the 19th century were more practically and scientifically oriented. Second, state institutions often invested most heavily on research in the dominant industries of the state (Goldin and Katz, 1999).

Figure 1 shows the location of colleges in counties over time. Figure 2 shows the variation in number of colleges across counties and over time. Figure 3 shows the variation in number of students across counties and over time.
Figure 1: Location of Colleges

(a) 1880

(b) 1930

Note: In each Panel, the location of a college is represented by a point on the map. A college is defined as existing if an observation occurs before or on the date of the graph and if there is an observation after or on the date of the graph. Source: All data comes from the Reports of the Commissioner of Education.
Figure 2: Geographic Distribution of Colleges

(a) 1880

(b) 1930

Note: In each Panel, counties are shaded according to their calculated number of colleges in 1880 (Panel A) and 1930 (Panel B). Counties are divided into five groups, and darker shades denote larger numbers of colleges. Source: All data comes from the Reports of the Commissioner of Education.
Figure 3: Geographic Distribution of Students

(a) 1880

(b) 1930

Note: In each Panel, counties are shaded according to their calculated number of students in 1880 (Panel A) and 1930 (Panel B). Counties are divided into six groups, darker shades denote larger numbers of students, and white represents missing data. Source: All data comes from the Reports of the Commissioner of Education.
Chapter 3

Data and Descriptive Statistics

Below I describe my data and sample selection.

3.1 Data Sources

My analysis uses several data sources which are linked by a county identifier or latitude and longitude data.

3.1.1 College Data

I use data from the Reports of the Commissioner of Education. This data is unique to my research. The data comprises all reporting colleges operating in the United States during the period from 1870 to 1940. I have data for every year annually from 1870 to 1913, 1915, and then every other year from 1916 until 1940. To the best of my knowledge, the next best available data sets would be reports translated for the years 1870, 1875, 1880, 1885, 1890, 1895, 1900, 1905, 1910, and 1914 by Heyu Xiong and Yiling Zhao and 1897, 1924, and 1934 by Claudia Goldin (Andrews, 2019). My dataset contains information on the name and location of the
college, date the college was organized, number of professors by gender, number of students by gender, religious affiliation, and the size of the university’s library. The location data was geocoded using opencagegeo (Zeigermann, 2018) and geocodehere (Heß, 2020) in 2019. I use a 1990 county shapefile created by Fabian Eckert, Andrés Gvirtz, Jack Liang, and Michael Peters (Eckert et al., 2018) which I then convert from Albers Equal Area Conic Projection to Latitude Longitude data as the county basemap. The college data was then assigned a county polygon from the resulting converted county shapefile.

### 3.1.2 Manufacturing Data

I use data from the US Census on Manufacturing which has been digitized for each county for the decennial census years 1870-1900 and 1920-1940. The dataset contains information from all manufacturing establishments with more than $500 in output (Walker, 1872). The manufacturing census includes data on number of establishments, number of employees, annual wages, costs of materials, level of capital invested in manufacturing, and value of output (Manson et al., 2016). In the initial productivity calculation I restrict the data used to value of manufacturing output and number of employees. The data for all years is adjusted to reflect 1990 counties (Eckert et al., 2018). All values in dollars are converted to 2018 dollars using the Consumer Price Index values reported for each year by the Minneapolis Federal Reserve Bank (FED, 2018).

### 3.1.3 Agricultural Data

I use Agricultural Census data which has been digitized for each county for the decennial census years 1870-1940. The dataset contains information on number
of farms, wages of farmers, costs of farm inputs, total amount of farmland, value of farmland, and agricultural output (Manson et al., 2016). In the initial productivity calculation I restrict the data used to agricultural output and total amount of farmland. The data for all years is adjusted to reflect 1990 counties (Eckert et al., 2018). All values in dollars are converted to 2018 dollars using the Consumer Price Index values reported for each year by the Minneapolis Federal Reserve Bank (FED, 2018).

### 3.1.4 Population Data

I use data on population from the decennial censuses which has been digitized for each county for the decennial years 1870-1940 (Manson et al., 2016).

### 3.1.5 Historic Railroad Data

I use data on railroads for the lower 48 states from 1776 through approximately 1911 with additional data through 1920. These files were created by Jeremy Atack using ESRI’s GIS software. Railroad locations come from the union of modern railroad SHP files published as a part of the National Atlas database by the U.S. Department of Transportation and USGS 7.5 minute topographical maps from the late nineteenth century onwards (Atack, 2018). The existence of a particular railroad at a given time was determined by its existence on maps of the time (Atack, 2018). These transportation files are the basis of most, if not all, research on the impacts of railroad expansion in the United States.
3.2 Sample Selection and Summary Statistics

Table 1 displays summary statistics for information on colleges. The table shows the means of number of professors, number of students, founding date, number of books in the library, and proportion of colleges that are religious, with standard deviations and percentage of observations missing in parenthesis. The number of students increases throughout the entire period.
## Table 1: Summary Statistics of College Variables

<table>
<thead>
<tr>
<th>Information On Professors</th>
<th>Years</th>
<th>1870-79</th>
<th>1880-89</th>
<th>1890-99</th>
<th>1900-09</th>
<th>1910-19</th>
<th>1920-29</th>
<th>1930-39</th>
<th>1940</th>
<th>1870-1940</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Professors</td>
<td>µ</td>
<td>10.076</td>
<td>10.658</td>
<td>14.531</td>
<td>22.964</td>
<td>39.843</td>
<td>83.465</td>
<td>406.786</td>
<td>240.954</td>
<td>76.328</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 7.549)</td>
<td>( 8.625)</td>
<td>( 20.405)</td>
<td>( 35.717)</td>
<td>( 105.552)</td>
<td>( 571.371)</td>
<td>( 2,437.809)</td>
<td>( 813.636)</td>
<td>( 847.582)</td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 7.061)</td>
<td>( 16.095)</td>
<td>( 0.348)</td>
<td>( 0.025)</td>
<td>( 0.099)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 3.055)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Male Professors</td>
<td>µ</td>
<td>12.398</td>
<td>18.108</td>
<td>32.700</td>
<td>69.134</td>
<td>309.981</td>
<td>181.892</td>
<td>73.571</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 19.974)</td>
<td>( 34.311)</td>
<td>( 102.827)</td>
<td>( 563.990)</td>
<td>( 1,950.161)</td>
<td>( 689.208)</td>
<td>( 776.288)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td>( 0.430)</td>
<td>( 1.368)</td>
<td>( 0.074)</td>
<td>( 0.518)</td>
<td>( 0.905)</td>
<td>( 24.232)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Female Professors</td>
<td>µ</td>
<td>2.151</td>
<td>4.936</td>
<td>7.166</td>
<td>15.521</td>
<td>119.696</td>
<td>68.440</td>
<td>22.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 4.116)</td>
<td>( 8.498)</td>
<td>( 13.526)</td>
<td>( 33.814)</td>
<td>( 1,308.708)</td>
<td>( 200.483)</td>
<td>( 454.112)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td>( 0.737)</td>
<td>( 1.231)</td>
<td>( 0.124)</td>
<td>( 5.895)</td>
<td>(12.945)</td>
<td>( 26.953)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information On Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Students</td>
<td>µ</td>
<td>99.901</td>
<td>103.235</td>
<td>136.776</td>
<td>219.849</td>
<td>386.967</td>
<td>875.333</td>
<td>1,262.723</td>
<td>1,436.622</td>
<td>435.844</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 115.729)</td>
<td>( 114.435)</td>
<td>( 218.026)</td>
<td>( 396.071)</td>
<td>( 716.578)</td>
<td>( 1,874.891)</td>
<td>( 2,898.995)</td>
<td>( 3,135.172)</td>
<td>( 1,383.777)</td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 10.665)</td>
<td>( 8.034)</td>
<td>( 0.778)</td>
<td>( 0.017)</td>
<td>( 0.000)</td>
<td>( 0.032)</td>
<td>( 0.000)</td>
<td>( 2.340)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Male Students</td>
<td>µ</td>
<td>87.992</td>
<td>77.858</td>
<td>101.516</td>
<td>141.576</td>
<td>246.109</td>
<td>577.819</td>
<td>935.375</td>
<td></td>
<td>305.642</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 103.552)</td>
<td>( 101.694)</td>
<td>( 182.072)</td>
<td>( 310.293)</td>
<td>( 526.172)</td>
<td>( 1,365.617)</td>
<td>( 2,278.396)</td>
<td>( 1,010.931)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 31.285)</td>
<td>( 37.620)</td>
<td>( 1.167)</td>
<td>( 0.274)</td>
<td>( 0.124)</td>
<td>( 3.947)</td>
<td>(12.518)</td>
<td>( 10.725)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Female Students</td>
<td>µ</td>
<td>28.466</td>
<td>18.806</td>
<td>35.753</td>
<td>79.109</td>
<td>141.407</td>
<td>333.995</td>
<td>535.906</td>
<td></td>
<td>174.607</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 47.966)</td>
<td>( 31.124)</td>
<td>( 75.952)</td>
<td>( 158.394)</td>
<td>( 302.976)</td>
<td>( 702.449)</td>
<td>(1,177.433)</td>
<td>( 1,249.238)</td>
<td>( 560.909)</td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 62.836)</td>
<td>( 51.549)</td>
<td>( 1.044)</td>
<td>( 0.615)</td>
<td>( 0.173)</td>
<td>( 4.995)</td>
<td>(15.066)</td>
<td>(16.312)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Founding Date</td>
<td>µ</td>
<td>1847.211</td>
<td>1850.834</td>
<td>1857.905</td>
<td>1860.137</td>
<td>1862.355</td>
<td>1868.788</td>
<td>1868.372</td>
<td></td>
<td>1867.585</td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td>( 0.000)</td>
<td></td>
</tr>
<tr>
<td>Number of Books in Library</td>
<td>µ</td>
<td>7,014.854</td>
<td>9,225.815</td>
<td>13,475.519</td>
<td>20,178.131</td>
<td>32,386.305</td>
<td>44,492.516</td>
<td>12,042.934</td>
<td></td>
<td>19,891.395</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(13,671.134)</td>
<td>(19,194.480)</td>
<td>(46,872.137)</td>
<td>(51,748.238)</td>
<td>(82,382.156)</td>
<td>(131,992.734)</td>
<td>(8,024.117)</td>
<td>(63,899.535)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 26.151)</td>
<td>( 69.617)</td>
<td>( 99.515)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td>( 100.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dummy Variable for Religious</td>
<td>µ</td>
<td>0.824</td>
<td>0.764</td>
<td>0.784</td>
<td>0.816</td>
<td>0.739</td>
<td>0.719</td>
<td>0.785</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>( 0.381)</td>
<td>( 0.424)</td>
<td>( 0.412)</td>
<td>( 0.387)</td>
<td>( 0.439)</td>
<td>( 0.450)</td>
<td>( 0.411)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>( 6.529)</td>
<td>(11.928)</td>
<td>(26.530)</td>
<td>( 0.889)</td>
<td>( 0.990)</td>
<td>( 84.040)</td>
<td>(100.000)</td>
<td>(100.000)</td>
<td>( 31.178)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td></td>
<td>3,385.000</td>
<td>3,647.000</td>
<td>4,885.000</td>
<td>5,850.000</td>
<td>4,042.000</td>
<td>4,054.000</td>
<td>3,093.000</td>
<td></td>
<td>29,659.000</td>
</tr>
<tr>
<td>Number of Unique IDs</td>
<td></td>
<td>529.000</td>
<td>457.000</td>
<td>694.000</td>
<td>725.000</td>
<td>794.000</td>
<td>995.000</td>
<td>864.000</td>
<td></td>
<td>1,473.000</td>
</tr>
</tbody>
</table>

**Note:** Dummy Variable for Religious represents proportion of colleges in period that are affiliated with a religion. Number of Unique IDs counts the number of different colleges that existed during the time period.

**Source:** All data comes from the Reports of the Commissioner of Education.
Table 2 displays summary statistics for key firm variables. The table gives means of the level of agricultural revenue, total land used for farming, agricultural revenue per acre, manufacturing revenue, number of employees, and manufacturing revenue per employee, with standard deviations and percent of observations missing in parenthesis. Agricultural revenue, total land used for farming, manufacturing revenue, and number of employees increase from 1870 to 1930 but fall in 1940 likely due to the Great Depression. Agricultural revenue per acre increases from 1880 to 1920 but then decreases in 1920 before increasing again. Agricultural revenue per acre in 1870 may be artificially high due to an incorrect approximation of farmland in 1870 described in section 5.3.3. By comparison, manufacturing revenue per employee increases throughout the entire period of study except for the year 1890.
Table 2: Summary Statistics of Firm Variables

<table>
<thead>
<tr>
<th>Information On Agriculture</th>
<th>Years</th>
<th>1870</th>
<th>1880</th>
<th>1890</th>
<th>1900</th>
<th>1910</th>
<th>1920</th>
<th>1930</th>
<th>1940</th>
<th>1870-1940</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of Agricultural Production (USD, thousands) µ</td>
<td>17,161</td>
<td>19,280</td>
<td>22,982</td>
<td>36,358</td>
<td>47,648</td>
<td>59,490</td>
<td>78,185</td>
<td>32,906</td>
<td>39,614</td>
<td>(22,984)</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(1)</td>
<td>(22,984)</td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(6)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(1)</td>
<td>(6)</td>
</tr>
<tr>
<td>Amount of Land Used for Agriculture (Acres, thousands) µ</td>
<td>45</td>
<td>179</td>
<td>209</td>
<td>261</td>
<td>284</td>
<td>307</td>
<td>635</td>
<td>341</td>
<td>286</td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(55)</td>
<td>(144)</td>
<td>(145)</td>
<td>(201)</td>
<td>(196)</td>
<td>(236)</td>
<td>(574)</td>
<td>(339)</td>
<td>(323)</td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(6)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(1)</td>
<td>(6)</td>
</tr>
<tr>
<td>Agricultural Production Per Acre (USD) µ</td>
<td>528</td>
<td>113</td>
<td>113</td>
<td>162</td>
<td>196</td>
<td>224</td>
<td>162</td>
<td>166</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(708)</td>
<td>(184)</td>
<td>(237)</td>
<td>(295)</td>
<td>(365)</td>
<td>(184)</td>
<td>(239)</td>
<td>(898)</td>
<td>(476)</td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(7)</td>
<td>(3)</td>
<td>(1)</td>
<td>(3)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(2)</td>
<td>(7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information On Manufacturing</th>
<th>Years</th>
<th>1870</th>
<th>1880</th>
<th>1890</th>
<th>1900</th>
<th>1910</th>
<th>1920</th>
<th>1930</th>
<th>1940</th>
<th>1870-1940</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of Manufacturing Production (USD, thousands) µ</td>
<td>28,192</td>
<td>47,937</td>
<td>89,867</td>
<td>129,086</td>
<td>250,752</td>
<td>753,279</td>
<td>391,878</td>
<td>226,610</td>
<td>(210,129)</td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(325,300)</td>
<td>(657,164)</td>
<td>(878,446)</td>
<td>(1,328,900)</td>
<td>(3,644,238)</td>
<td>(1,943,002)</td>
<td>(1,644,207)</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(1)</td>
<td>(0)</td>
<td>(4)</td>
<td>(3)</td>
<td>(100)</td>
<td>(5)</td>
<td>(16)</td>
<td>(19)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>Total Number of People Employed in Manufacturing µ</td>
<td>689</td>
<td>944</td>
<td>1,617</td>
<td>1,746</td>
<td>3,079</td>
<td>6,738</td>
<td>2,861</td>
<td>2,436</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(4,148)</td>
<td>(6,190)</td>
<td>(10,073)</td>
<td>(10,318)</td>
<td>(16,661)</td>
<td>(33,578)</td>
<td>(14,563)</td>
<td>(15,993)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(1)</td>
<td>(0)</td>
<td>(4)</td>
<td>(3)</td>
<td>(100)</td>
<td>(5)</td>
<td>(16)</td>
<td>(11)</td>
<td>(18)</td>
<td></td>
</tr>
<tr>
<td>Manufacturing Production Per Employee (USD, thousands) µ</td>
<td>36</td>
<td>55</td>
<td>45</td>
<td>80</td>
<td>90</td>
<td>90</td>
<td>242</td>
<td>76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(24)</td>
<td>(47)</td>
<td>(20)</td>
<td>(49)</td>
<td>(76)</td>
<td>(90)</td>
<td>(90)</td>
<td>(69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Percent Missing)</td>
<td>(10)</td>
<td>(9)</td>
<td>(0)</td>
<td>(0)</td>
<td>(100)</td>
<td>(5)</td>
<td>(16)</td>
<td>(24)</td>
<td>(22)</td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>3,019</td>
<td>3,034</td>
<td>3,048</td>
<td>3,111</td>
<td>3,109</td>
<td>3,111</td>
<td>3,111</td>
<td>3,111</td>
<td>24,654</td>
<td></td>
</tr>
</tbody>
</table>

Source: Data on agriculture comes from the Census of Agriculture and data on manufacturing comes from Census of Manufacturers both of which are published by the Census Bureau and made available by IPUMS NHGIS.
Figure 4 shows the variation in agricultural output per acre across counties and over time. Figure 5 shows the variation in manufacturing output per employee across counties and over time.
Figure 4: Geographic Distribution of Agricultural Output Per Acre (USD)

(a) 1880

(b) 1930

Note: In each Panel, counties are shaded according to their calculated level of agricultural output per acre in 1880 (Panel A) and 1930 (Panel B). Counties are divided into six groups, darker shades denote higher agricultural output per acre, and white represents missing data. Source: All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Figure 5: Geographic Distribution of Manufacturing Output Per Employee (USD)

(a) 1880

(b) 1930

Note: In each Panel, counties are shaded according to their calculated level of manufacturing output per employee in 1880 (Panel A) and 1930 (Panel B). Counties are divided into six groups, darker shades denote higher manufacturing output per employee, and white represents missing data. Source: All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Chapter 4

Empirical Model

4.1 Defining Productivity

My two main outcomes of interest are agricultural productivity and manufacturing productivity. I define agricultural productivity ($PA_{ct}$) as agricultural revenue per acre. I define manufacturing productivity ($PM_{ct}$) as manufacturing revenue per employee:

\begin{align*}
PA_{ct} &= \frac{AR_{ct}}{A_{ct}} \quad (1) \\
PM_{ct} &= \frac{MR_{ct}}{L_{ct}} \quad (2)
\end{align*}

where $AR_{ct}$ is the value of agricultural output in county $c$ and year $t$, $A_{ct}$ is the total number of acres used for farming in county $c$ and year $t$, $MR_{ct}$ is the value of manufacturing output in county $c$ and year $t$, and $L_{ct}$ is the total number of employees employed by all manufacturing firms\footnote{Firms were instructed to report the average number of employees they employed during the year but sometimes reported the number of employees they employed on the day of data collection.} in county $c$ and year $t$. I follow Kantor and Whalley (2019) in regarding agricultural revenue per acre as a measure of the revenue
productivity of land.

4.2 Estimating Equation

I regress outcome, y, where y represents either PA_{ct} or PM_{ct}, in county c and time t on the average number of students enrolled in universities since the last census, population, county fixed effects, and county-by-year fixed effects.

\[
y_{ct} = \beta_0 + \beta_1 \text{students}_{ct} + \beta_2 \text{population}_{ct} + \gamma_c + \gamma_{ct} + \varepsilon_{ct}
\]  

(3)

The coefficient \( \beta_1 \) reports the effect of average number of students on outcome, y, comparing changes in counties with relative increases in average number of students to other counties controlling for population, county fixed effects, and county-by-year fixed effects. The identification assumption is that counties with relative increases in number of students would otherwise have changed similarly to nearby counties.
Chapter 5

Empirical Results

5.1 Estimated Impacts on Productivity

I use equation (3) to estimate the effect of colleges on agricultural and manufacturing productivity. Table 3 presents the effects of an increase in the average number of students enrolled in universities since the last census on agricultural and manufacturing productivity, respectively. Panel 1 reports the impact of average number of students on agricultural and manufacturing productivity controlling for county and county-by-year fixed effects only. Panel 2 reports the impact of average number of students on agricultural and manufacturing productivity adding a control for population.

The results show that an increase in the average number of students enrolled in universities increases productivity in the agricultural and manufacturing sectors. Although these estimates need to be interpreted with caution, I find that every one student increase in the average number of students enrolled in a university in a county over the preceding decade increases manufacturing output per employee by $2.13, holding all else equal. This is significant at the 95% significance level. Additionally,
**Panel 1. Productivity Growth. Raw**

<table>
<thead>
<tr>
<th></th>
<th>Agricultural</th>
<th>Manufacturing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>0.014</td>
<td>2.237</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.006)</td>
<td>(0.986)</td>
</tr>
<tr>
<td>Adj. (R^2)</td>
<td>0.509</td>
<td>0.698</td>
</tr>
<tr>
<td>N</td>
<td>2973.000</td>
<td>2430.000</td>
</tr>
<tr>
<td>Outcome Mean</td>
<td>212.265</td>
<td>84367.133</td>
</tr>
</tbody>
</table>

**Panel 2. Productivity Growth. Control for Population**

<table>
<thead>
<tr>
<th></th>
<th>Agricultural</th>
<th>Manufacturing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>0.013</td>
<td>2.131</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.006)</td>
<td>(0.990)</td>
</tr>
<tr>
<td>Adj. (R^2)</td>
<td>0.510</td>
<td>0.698</td>
</tr>
<tr>
<td>N</td>
<td>2973.000</td>
<td>2430.000</td>
</tr>
<tr>
<td>Outcome Mean</td>
<td>212.265</td>
<td>84367.133</td>
</tr>
</tbody>
</table>

\[1 \ y_{ct} = \beta_0 + \beta_1 \text{students}_{ct} + \gamma_c + \gamma_{ct} + \varepsilon_{ct} \]

\[2 \ y_{ct} = \beta_0 + \beta_1 \text{students}_{ct} + \beta_2 \text{population}_{ct} + \gamma_c + \gamma_{ct} + \varepsilon_{ct} \]

I find that every one student increase in the average number of students enrolled in a university in a county over the preceding decade, increases agricultural output per acre by $.01, holding all else equal. This effect is modest but significant at the 95% confidence level.

Figure 6 graphs mean agricultural output per acre (a) and manufacturing output per employee (b) for counties with a university and counties without a university by year. Figure 6a shows that agricultural productivity is higher in counties without a college in 1870 but counties with a college became more productive starting in 1880. Panel 6b shows that manufacturing productivity was about equal in counties with and without a university in 1870. This effect persists until 1920 when counties with a college became consistently more productive on average.
Figure 6: Actual and Counterfactual Productivity Measures

(a) Agricultural Revenue Per Acre (USD)

(b) Manufacturing Revenue Per Employee (USD)

Note: each panel compares mean agricultural output per acre (Panel A) and mean manufacturing output per employee (Panel B) for counties with a college and counties without a college at the decennial years from 1870 to 1940. Source: Data on colleges comes from the Reports of the Commissioner of Education. Data on agriculture comes from the Census of Agriculture and data on manufacturing comes from Census of Manufacturers both of which are published by the Census Bureau and made available by IPUMS NHGIS.
5.2 Theoretical Explanation

Colleges have been hypothesized to increase productivity through two main channels: direct effects that involve the direct interaction between faculty, students, and firms and indirect effects such as spillovers and agglomeration economies. Colleges directly increase human capital through three mechanisms. First, universities facilitate interaction between faculty and local business establishments (Liu, 2015). Second, they increase the local supply of human capital by training students who remain in the area (Abel and Deitz, 2012). Third, faculty conducts research which develops new ideas or processes (Andrews, 2019). It has been shown that human capital increases individual-level productivity and idea generation (Becker, 1993). Therefore, we would expect that a higher level of human capital within a region would increase regional productivity. Education is of great value in agriculture since it helps farmers adapt more quickly to new hybrids and other new technologies (Becker, 1993). Colleges may also increase productivity through indirect channels. The most prominent of these mechanisms is that colleges increase the population of the surrounding region leading to agglomeration economies. Agglomeration economies can increase productivity through three channels: intermediate input sharing, knowledge spillovers, and labor market pooling (Marshall, 1890). My study attempts to control for the effect of agglomeration economies by including a control for population.

5.3 Measurement Error

The measured county productivity may be subject to various types of measurement error.
5.3.1 College Error

During the process of converting the data to a usable format I had to drop 1784 observations making up 5.67% of the dataset. These observations were dropped for various reasons. Forty-three duplicate observations were dropped. Eighty-four observations were dropped because the college name was ambiguous, likely contained an error, appeared to summarize a category as opposed to being an observation for a college, or the college was part of a US territory that has not become a state. Twenty-seven observations were dropped because they were located in Hawaii or Alaska which I did not have county data for. Four hundred thirty-six observations were dropped because I was unable to assign them geographic coordinates. One thousand one hundred ninety-four observations were dropped because I was unable to assign the college a founding date. Additionally, I homogenized the locations of colleges to one location to reduce the chance that the same college would be geocoded to multiple addresses. However, I tried to maintain the location for an observation if it was obvious the college had moved over time. This process may have resulted in a college being placed in the incorrect county if the college had moved and I incorrectly assumed it had not.

5.3.2 Manufacturing Data

The data reported for manufacturers is susceptible to several sources of error. First, census enumerators were paid on a per-firm basis meaning in rural areas which are less densely populated census enumerators had less of an incentive to survey isolated manufacturing activities (Atack and Bateman, 1999). However, a review of the census manuscripts by Jeremy Atack “reveal[s] evidence of the care and attention to detail paid by so many enumerators” (Atack and Bateman, 1999). Second, the
census data may have been manipulated by “replacing information reported by the original enumerators with other (more appropriate?) numbers” (Atack and Bateman, 1999) by compilers in Washington D.C. Third, the tallies recorded were incorrect and inconsistent with the reports by enumerators. Fourth, like the college data there may have been mistakes in translating from census reports of the time to a usable dataset. Additionally, county level manufacturing data was not collected for 1910 so I am unable to calculate manufacturing productivity in 1910.

5.3.3 Agricultural Data

The agricultural data suffers from two major pitfalls. First, the total amount of farmland in a county was not reported for the years 1870, 1880, and 1890. I, therefore, had to estimate total farmland for these years. For the years 1880 and 1890, average farm size was reported. I calculated total farm land in a county by multiplying the average size of the farm by the number of farms. In 1870, the census provided a breakdown of the number of farms belonging to nine size categories\(^1\). I estimated the amount of farmland in 1870 by multiplying the number of farms belonging to a category by the lower bin size. For example, if the category was 10-19 acres and contained 10 farms I calculated the total farm size for that group to be 100 acres. This was the specification that minimized the mean difference between the estimated farm land amount and the true amount of farmland in a county in 1900 and the calculated level for 1880 and 1890. Second, agricultural output was reported as total farm revenue from 1870-1900 and then total revenue from crop production not fed to livestock from 1910-1940. This change in specification decreased output for the years 1910-1940 relative to the pre-1910 period. I control for this by including

\(^1\)less than 3 acres, 3-9 acres, 10-19 acres, 20-49 acres, 50-99 acres, 100-499 acres, 500-999 acres, and 1000 acres or more
county-by-year fixed effects.

5.4 Endogeneity Problem

The main challenge of estimating the effect that universities have on labor productivity is that university activity is not random. This endogeneity occurs because the universities are possibly affected by the existence of highly productive firms. This impacts the college in several ways. First, highly productive regions are likely to be wealthier and, therefore, possess the political capital necessary to lobby for a college. Second, these regions are likely to have higher demand for education either by increased demand for an educated labor force or parents of wealthy families wanting to educate their children. Third, highly productive individuals may provide human or physical capital required by the college to be successful.

Several strategies have been proposed to counteract this problem. The first and most popular is to assume the locations of colleges are random (Currie and Moretti (2003), Moretti (2004), Liu (2015)). These studies typically look specifically at land-grant colleges in the US. An example of this is Moretti’s Estimating the Social Return to Higher Education: Evidence From Longitudinal and Repeated Cross-Sectional Data which states that “Land-grant colleges were often established in rural areas, and their location was not dependent on natural resources or other factors that could make an area wealthier. In fact, judged from today’s point of view, the geographical location of land-grant colleges seems close to random” (Moretti, 2004). The second method is using synthetic control counties to run a difference-in-difference regression. This method uses weighted averages of counties where the weights are chosen to ensure the synthetic county closely matches the treated county (Liu, 2015). The third method is to use instrumental variables. Some instruments that have been
used in the literature are stock market shocks on endowment spending (Kantor and Whalley, 2014) and establishment of agricultural experiment stations at preexisting land-grant colleges (Kantor and Whalley, 2019). Fourth, is using narrative history to identify runner-up sites and universities whose location is as good as random and running a difference-in-difference regression. This methodology relies on the idea that when selecting where to locate a new college, dozens of possible candidate locations are considered and iteratively eliminated; by the time only a few finalists sites are left they are likely similar along both observable and unobservable dimensions (Andrews, 2019). I counteract the endogeneity problem by including county and county-by-year fixed effects. I assume colleges are plausibly located randomly at the county level when controlling for county and county-by-year.

5.5 Robustness Checks

I conduct additional analysis to evaluate several possible concerns.

5.5.1 Restricting the Time Period

First, I restrict the period of study to the timeframe 1870-1930. The Great Depression caused massive changes in the economic landscape of the United States. We would expect that levels of output would fall substantially during this time. This specification leads to the effect of lagged number of students on agricultural productivity becoming insignificant but manufacturing productivity remains significant although with a slightly reduced effect.\footnote{\$1.67 vs \$2.13 per employee}
5.5.2 Regression on Current Number of Students

Second, I estimated the impact of current number of students on my productivity estimates. Agricultural productivity is significant at the 99% significance level and the point estimate is left unchanged. Manufacturing productivity is significant at the 90% level and the point estimate is reduced to $1.49 per employee.

5.5.3 Regression on Average Number of Students Since the Last Census Per Capita

Third, I estimate the impact of average number of students since the last census per capita on my productivity estimates. This specification results in point estimates that are insignificant for agricultural productivity and manufacturing productivity.

5.5.4 Alternative Total Farmland Assumption for 1870

One concern is that my estimation method for the total amount of farmland in 1870 underestimates the amount of farmland. Therefore, I estimated the amount of farmland in 1870 by multiplying the number of farms belonging to a category by the midpoint of the bin size instead of the lower bound. For example, if the category was 10-19 acres and contained 10 farms, I calculated the total farm size for that group to be 145 acres. This specification results in point estimates that are insignificant for agricultural productivity.

5.5.5 Standardize Farm Revenues

Another concern is that the county-by-year fixed effects do not control for the changed method of reporting agricultural revenue between 1900 and 1910. I
standardize farm revenue to a within-year mean of zero and standard deviation of one, using the year mean and standard deviation. This specification results in point estimates that are insignificant for agricultural productivity.

5.5.6 Alternate Productivity Measure

Another potential measure of productivity growth is to calculate the growth in producer surplus:

\[
PA_{ct} = dAR_{ct} - dAC_{ct} \tag{4}
\]
\[
PM_{ct} = dMR_{ct} - dMC_{ct} \tag{5}
\]

where \( dAR_{ct} \) is growth in the total value of agricultural output in county \( c \) and year \( t \), \( dAC_{ct} \) is growth in the total value of inputs used in agriculture in county \( c \) and year \( t \), \( dMR_{ct} \) is growth in the total value of manufacturing output in county \( c \) and year \( t \), and \( dMC_{ct} \) is growth in the total value of inputs used in manufacturing in county \( c \) and year \( t \). A variation of this specification is referred to as productivity growth by Solow, 1957; Basu and Fernald, 2002; Petrin and Levinsohn, 2012; Baqee and Farhi, 2017; Donaldson and Hornbeck, 2019. This specification results in point estimates that are insignificant for agricultural productivity and manufacturing productivity. This method may suffer from significant measurement error in the methods by which inputs are calculated. For example, the types of costs reported for farms vary from census to census\(^3\) (Manson et al., 2016) and costs for labor and capital may be under-

\(^3\)In 1870, the census reports the expenditures on wages and implements and machinery. In 1880, the census reports expenditures on fertilizer, fencing, and implements and machinery. In 1890, the census reports expenditures on fertilizer and implements and machinery. In 1900, the census reports expenditures on fertilizer, wages, irrigation, and implements and machinery. In 1910, the census reports expenditures on wages, rent, fertilizer, feed, irrigation, and implements and machinery. In 1920, the census reports expenditures on wages, feed, fertilizer, and implements and machinery. In 1930, the census reports expenditures on feed, fertilizer, labor, machinery, power. In 1940, the
5.5.7 Include Control for Distance to Railroad

According to Hornbeck and Rotemberg (2019) and Fogel (1964) railroads were a significant innovation in the transportation sector that generated large economic gains. I include distance to a railroad as an additional control. The regression equation then becomes:

\[ y_{ct} = \beta_0 + \beta_1 students_{ct} + \beta_2 population_{ct} + \beta_3 distrailroad_{ct} + \gamma_c + \gamma_t + \varepsilon_{ct} \]  

(6)

A pitfall of this method is that GIS railroad data is only available up to 1920 and, therefore, I had to restrict the period of study to 1920. This specification results in estimates that are insignificant for manufacturing productivity. However, agricultural productivity is significant at the 95% confidence level and the point estimate increases from $.01 to $.05.

census reports expenditures on implements and machinery only

30
Chapter 6

Conclusion

The period following 1870 shaped the modern research university in the United States. The number of individuals enrolled in universities and the economy of the United States increased dramatically during the period from 1870 to 1940.

I find that every one student increase in the average number of students enrolled in a university in a county over the previous decade increases agricultural output per acre by $.01 and manufacturing output per employee by $2.13. Additionally, I find that the effect of universities on agricultural productivity occurs as early as 1880 but the effect on manufacturing productivity takes longer to manifest and does not appear significant until 1920.

6.1 Limitations Of Study

It is possible that these results are undermined if colleges are not located randomly and are instead influenced by the presence of highly productive firms.

Additionally, lack of consistent data throughout the period of study limited the types of studies that could be conducted and required me to make possibly inaccurate
assumptions that may undermine my findings.

6.2 Recommendations for Further Research

My study provides three directions for future work. First, developing a more sophisticated estimation strategy such as a synthetic control method to better control for the concern that the establishment and effectiveness of colleges are not random. Second, more comprehensively compensating for inaccurate data in the Census of Manufacturers and Agriculture to calculate productivity growth using change in producer surplus as a metric. Third, extending the work of Jeremy Atack on historic transportation networks through 1940 so that market access can be included as a control for the entire period of study.
Bibliography


Heß, S. (2020, February). GEOCODEHERE: Stata Module to Provide Geocoding Relying on Nokia’s Here Maps API. original-date: 2016-06-13T17:00:35Z.


Manson, S., J. Schroeder, D. Van Riper, and S. Ruggles (2016, July). Citation and Use of NHGIS Data.


Appendices
Appendix A  Location Maps
Figure 7: Location of Colleges in the U.S.

(a) 1870

(b) 1880
Note: In each Panel, the location of a college is represented by a point on the map. A college is defined as existing if an observation occurs before or on the date of the graph and if there is an observation after or on the date of the graph. Source: All data comes from the Reports of the Commissioner of Education.
Figure 8: Geographic Distribution of Railroads

(a) 1870

(b) 1880
Note: In each Panel, the location of a railroad is represented by a line on the map. A railroad is defined as existing by Jeremy Atack if it was shown on maps of the time. **Source:** All data comes from Jeremy Atack’s work on historic transportation shape files.
Figure 9: Geographic Distribution of Navigable Rivers

(a) 1870

(b) 1880
Note: In each Panel, the location of a river is represented by a line on the map. A river is defined as existing by Jeremy Atack if it was shown on maps of the time. Source: All data comes from Jeremy Atack’s work on historic transportation shape files.
Figure 10: Geographic Distribution of Canals

(a) 1870

(b) 1880
Note: In each Panel, the location of a canal is represented by a line on the map. A canal is defined as existing by Jeremy Atack if it was shown on maps of the time. Source: All data comes from Jeremy Atack’s work on historic transportation shape files.
Figure 11: College Locations with Access to Transportation Networks

(a) 1870

(b) 1880
Note: In each Panel, the location of colleges are represented by points on the map, railroads are represented by a grey line, and rivers are located by a blue line. A college is defined as existing if an observation occurs before or on the date of the graph and if there is an observation after or on the date of the graph. Railroads and rivers are defined as existing by Jeremy Atack if they were shown on maps of the time. Source: Data on colleges comes from the Reports of the Commissioner of Education. Data on transportation networks comes from Jeremy Atack’s work on historic transportation shape files.
Figure 12: Geographic Distribution of Colleges

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of colleges in each year. Counties are divided into five groups, and darker shades denote larger numbers of colleges. Source: All data comes from the Reports of the Commissioner of Education.
Figure 13: Geographic Distribution of College Openings

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of colleges opened during the previous decade. Counties are divided into five groups, and darker shades denote larger numbers of college openings. Source: All data comes from the Reports of the Commissioner of Education.
Figure 14: Geographic Distribution of Students

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of students in each year. Counties are divided into six groups, darker shades denote larger numbers of students, and white represents missing data. Source: All data comes from the Reports of the Commissioner of Education.
Figure 15: Geographic Distribution of Professors

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of professors in each year. Counties are divided into six groups, darker shades denote larger numbers of professors, and white represents missing data. Source: All data comes from the Reports of the Commissioner of Education.
Appendix B  Maps of Agricultural Statistics
Figure 16: Geographic Distribution of Farms

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of farms in each year. Counties are divided into six groups, darker shades denote higher numbers of farms, and white represents missing data. Source: All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Figure 17: Geographic Distribution of Wages Paid to Farmers (USD)

(a) 1870

(b) 1900
(c) 1910

(d) 1920
Note: In each Panel, counties are shaded according to their calculated level of wages paid to farmers in each year. Counties are divided into six groups, darker shades denote higher levels of wages paid to farmers, and white represents missing data. Source: All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Figure 18: Geographic Distribution of Agricultural Revenue (USD)

(a) 1870

(b) 1880
(e) 1910

(f) 1920
Note: In each Panel, counties are shaded according to their calculated level of agricultural output. Counties are divided into six groups, darker shades denote higher levels of agricultural output, and white represents missing data. **Source:** All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Figure 19: Geographic Distribution of Farm Land (Acres)

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated amount of farmland. Counties are divided into six groups, darker shades denote higher amounts of farmland, and white represents missing data. Source: All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Figure 20: Geographic Distribution of Agricultural Revenue Per Acre (USD)

(a) 1870

(b) 1880
(e) 1910

(f) 1920
Note: In each Panel, counties are shaded according to their calculated level of agricultural output per acre. Counties are divided into six groups, darker shades denote higher levels of agricultural output per acre, and white represents missing data. Source: All data comes from the Census of Agriculture published by the Census Bureau and made available by IPUMS NHGIS.
Appendix C  Map of Manufacturing Statistics
Figure 21: Geographic Distribution of Manufacturing Establishments

(a) 1870

(b) 1880

Legend:

- 0
- <=10
- 10-28
- 28-73
- >73
- Missing
Note: In each Panel, counties are shaded according to their calculated number of manufacturing firms in each year. Counties are divided into six groups, darker shades denote higher numbers of manufacturing firms, and white represents missing data. Source: All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Figure 22: Geographic Distribution of Wages Paid to Manufacturing Employees (USD)

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated level of wages paid to manufacturing workers in each year. Counties are divided into six groups, darker shades denote higher levels of wages paid to manufacturing workers, and white represents missing data. **Source:** All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Figure 23: Geographic Distribution of Revenue From Manufacturing Activity (USD)

(a) 1870

(b) 1880

Legend:
- 0
- <=2100000
- 2100000-10000000
- 10000000-51000000
- >51000000
- Missing
Note: In each Panel, counties are shaded according to their calculated level of manufacturing output. Counties are divided into six groups, darker shades denote higher levels of manufacturing output, and white represents missing data. **Source:** All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Figure 24: Geographic Distribution of Manufacturing Employees

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated number of manufacturing employees. Counties are divided into six groups, darker shades denote higher numbers of manufacturing employees, and white represents missing data. **Source:** All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Figure 25: Geographic Distribution of Manufacturing Revenue Per Employees (USD)

(a) 1870

(b) 1880
Note: In each Panel, counties are shaded according to their calculated level of manufacturing output per employee. Counties are divided into six groups, darker shades denote higher levels of manufacturing output per employee, and white represents missing data. Source: All data comes from the Census of Manufacturers published by the Census Bureau and made available by IPUMS NHGIS.
Appendix D    Additional Graphs
Figure 26: Outcome Variable and Lagged Number of Students

(a) Agricultural Revenue Per Acre (USD) and Lagged Number of Students

Note: each panel compares mean agricultural output per acre (Panel A) and mean manufacturing output per employee (Panel B) for counties with a college to the mean lagged number of students at the decennial years from 1870 to 1940. Source: Data on colleges comes from the Reports of the Commissioner of Education. Data on agriculture comes from the Census of Agriculture and data on manufacturing comes from Census of Manufacturers both of which are published by the Census Bureau and made available by IPUMS NHGIS.