Sectarian Competition and the Market Provision of Human Capital

PRELIMINARY AND INCOMPLETE

Heyu Xiong*†
Yiling Zhao‡

June, 2019

Abstract

In the latter half of the 19th century, America experienced a significant expansion in its collegiate infrastructure. By 1890, more institutions of higher learning existed in the United States than all of Europe. In this paper we study the role of denominational competition in the market provision of higher education. Specifically, we document nearly all colleges established in this time period had denominational roots or origins. The empirical analysis reveals a robust positive relationship between an area's religious fragmentation and the number of colleges established locally. We argue that denominational affiliation facilitated enthusiasm to build colleges through gains to differentiation from standard Hotelling channels. We formulate a model of school choice, entry, and denominational affiliation. We find evidence that differentiation softened the extent of tuition competition and mediated an "excess" entry of colleges. We conclude by showing that the higher equilibrium quantity of schools, associated with increased entry, had persistent effect on institutional quality; thus, religious diversity precipitated educational investment.

*We thank Joel Mokyr, Carola Frydman, Nancy Qian, and Joseph Ferrie for their thoughts and comments on this work. We also thank economic history lunch seminar participants at Northwestern University. We thank Ashish Agarwal for his research assistance. All remaining errors are our own. The authors are grateful for financial support financial support from the Balzan Foundation and Northwestern's Center for Economic history.

†Department of Economics, Northwestern University, E-mail: heyuxiong2018@u.northwestern.edu.
‡Department of Economics, Northwestern University, E-mail: yilingzhao2018@u.northwestern.edu.

1 Introduction

In the United States, the closing decades of the 19th century witnessed a significant expansion in the scope and landscape of higher learning. From 1820 to 1859, 225 private universities were built; and over the next 40 years, an additional 348 were founded. By 1881, Ohio alone had 43 such institutions to minister to its 3 million inhabitants. In contrast, England had merely four universities serving a population of 23 million Goetzmann (2009). The high density of private universities and their remarkable quality would become an enduring characteristic of the American higher education system. But where did this growth in educational infrastructure originate? And what sustained such seemingly disproportionate level of investment?

This article studies the economic and competitive forces which shaped the zealous pattern of American collegiate development on the eve of its *formative* years. In particular, we highlight the role of denominational competition in the market provision of higher education prior to 1890. We document that virtually all private colleges established in this period had denominational roots or origins. This sectarian ethos of early college entry reflected the religious tenor of 19th century American life. Owing to the diversity of religious composition in American population, denominational affiliation was an instrument of strategic choice and a source of product differentiation. We argue that differentiation along the dimension of religion lessened the degree of competition amongst colleges and mediated an "excess" of entrants within narrow geographical markets.

Scholars have long noted the "virtues" of the US education system within the first hundred years of the republic's inception Goldin and Katz (2010). A large strand of literature has emphasized the fundamental principles of fiscal independence, secular control, and public provision, crucial to its success and diffusion. However, with notable exceptions, this existing literature has focused on elementary and secondary schooling which constituted the bulk of mass education. Largely omitted is discussions of higher education prior to 1890; where, paradoxically, denominational and private interests played a substantial and more nuanced role. By exploring the origins and foundations of the "knowledge industry", this paper contributes a missing but complementary chapter to the story of early American educational "exceptionalism".

At a glance, the significance of colleges in the mid nineteenth century is easy to overlook. During the antebellum and postbellum years, the fraction of college educated persons in the population was, as they were in all nations, small. Yet the mere number of degrees conferred belies the significance of universities in the later stages of industrialization, a key historical period referred to as the Second Industrial Revolution. There is increasing evidence that knowledge at the upper tails mattered significantly more for economic development than average human cap-

ital or literacy (Squicciarini and Voigtländer, 2014) and universities themselves played startling role in facilitating the Commercial and Industrial revolutions (Cantoni and Yuchtman, 2012; Mokyr, 2009). Furthermore, Card (2001) documents that even proximity to colleges impact decisions of educational attainment.[1] The striking implication of Card (2001) is that the equilibrium distribution of universities have a first order impact on the aggregate educational stock of a country's labor force. If the 20th century is designated the "human-capital century", this paper endeavors to analyze the initial conditions to which that is owed.

We begin by compiling, from several primary and secondary sources, the universe of all known bachelor granting institutions in 19th century US. For each institution, we document the exact geographic location, land-grant status, the cost of tuition and board, the resources provided, the level of endowment, and, importantly, the denominational affiliation if stated. Using the geographic information, we link these institutions back to the underlying markets in which they were initially built.

Combining decadal population censuses with the censuses of religious bodies covering the same period, we assemble a rich panel dataset of US counties spanning 1850 to 1890 with detailed accounts of religious composition and memberships. Specifically, we collect data on the number of churches belonging to each denominations within the county and their respective market share. From this information, we compute a measure of a county's religious market concentration or fractionalization.

The first part of our empirical analysis reveals a significant positive relationship between religious diversity and college establishments, both in the cross section and over time. We relate the changes in number of colleges to changes in levels of religious fragmentation. A back-of-the-envelope calculation using our panel estimates suggests that there would have been approximately 20 percent fewer colleges established between 1850 and 1890 if U.S was dominated by a single denomination. We interpret this correlation as evidence for the role of religious competition in spurring college provision. The estimated coefficient is robust to the inclusion of an exhaustive set of confounding factors. We control for county fixed effects, denomination fixed effect, population, urbanization, industrialization, education, gender ratio, land productivity, agricultural output, geographic conditions, and railway or canal access.

By controlling for the flows of migrants, we exploit differences in the religious composition of the population to identify the religious competition channel, holding constant county characteristics that attract overall migration rate regardless of their religious origins. The denomination fixed effects imply the positive association reflects the effect of religious fragmentation as op-

[1] In a concurrent project, Andrews (2017) examines the causal effect of colleges on innovation in the 19th century.

posed to the presence of any single or combinations of denominations. The addition of county fixed effect accounts for unobserved county heterogeneity and limits our identifying variation to only within county changes to religious composition over time. This mitigates issues of selection.

To rule out competing explanations, we conduct falsification checks. One may think that religious diversity is possibly correlated with unobserved preferences for education, and thereby college building is an expression of this broader omitted variable. Placebo analysis show religious diversity had no impact on the number of high schools or land grant colleges. The relationship only exists between religious competition and the supply of, specifically, private denominational colleges.

There is possibly concerns of reverse causality if counties with high density of universities are areas exhibiting high economic potential and attract migratory flows from diverse background. To address this, we analyze whether 1860 college concentration predicted religious diversity in 1890 and it did not. Religious diversity explains college entry and not the other way around.

To complement the U.S.-wide analysis, our second empirical approach uses an instrumental variable strategy based on a particular historical episode observed in a sub-sample of data. While narrower in scope, the research design is arguably more credible. Specifically, we identify plausibly exogenous variation in religious competition induced by the onset and diffusion of the Second Great Awakening in the 1830s. During this period, the United States underwent a religious resurgence and unprecedented expansion of church memberships. Emphasizing the democratization of Protestantism, the Second Great Awakening leveled church hierarchies and gave arise to new religious sects.

This period of religious fervor and enthusiasm was characterized by frequent religious gatherings, termed revivals, which facilitated frenzied mass conversions, personal religious experiences which caused individuals to devote themselves religiously. Because participation in revivals differed between denominations and new denominations birthed from these events, we consider these spiritual revivals as quasi-natural experiments which fundamentally altered the religious fabric in a given locality. We provide evidence, using township level data from New York State, that revival events increased local religious competition and led to development of higher education.

An extensive literature, beginning with the pioneering work of Max Weber, have explored the connection between religious *norms* and educational or economic outcomes. Our paper contributes to this literature by illustrating that competition and interaction *between* narrowly defined denominations can impact equilibrium provision of upper-tail human capital. Compar-

atively, there is relatively little empirical research on the consequences of religious diversity and the existing works on religion have largely offered channels that are sociological rather than economic.[2]

Drawing from the industrial organization literature on differentiated product and entry, we attempt to disentangle the underlying mechanisms specific to our context. Starting a college in 19th century US was a precarious proposition which carried significant risk of failure. In this competitive environment, we posit that denominational differentiation provided a mechanism to insulate entrants from onerous competition on tuitions and to extract higher rent.

We build on our reduced form evidence and rationalize the results within the framework of historical university competition. We consider the college denomination to be a dimension for horizontal differentiation so that colleges can cater to specific consumer preferences via affiliation choices. Denominational differences effectively rendered colleges less substitutable. The key results hinge on two crucial assumptions: one, students exhibit preferences for colleges whose denominational affiliation match their own; two, colleges recruit students locally.

Without fully formalizing the model, we assess the strength of these parameters which govern the underlying incentives and gains to differentiation. We use individual micro data to estimate an empirical model of school choice and assess how student preferences vary in the population. To our knowledge, little is known regarding the demand side of colleges in the 19th century. We construct a unique student level dataset, assembled from college directories, alumni records and biographies, which contains information on student characteristics as well as the college attended. We estimate a conditional logit discrete choice model and diagnose how the various factors such as religious affiliation, distance, and tuition affected observed choice.

We find students strongly preferred to attend college with denomination that matched their own. Our preferred estimates indicate that students were willing to travel up to 250 kilometers further and pay an additional $500 per semester for a college with that quality. The revealed preferences imply that colleges, even in close proximity, can secure sufficient demand by differentiating themselves denominationally. And consequently, the number of entrants sustained in equilibrium will be increasing in the denominational heterogeneity in the underlying population. In a counterfactual world with less local religious diversity or mandated secular universities in exclusion of religious interests, the number of colleges established would be substantially lower.

The strength of the U.S. higher education system lies is in both the quantity as well as the quality of the schools. In the last part of the paper, having shown that religious diversity led to

[2]The broad interaction of Catholicism and Protestantism is known to shape the primary education provision in Europe (Stone, 1968).

increased entry and a more competitive environment, we investigate the relationship between market structure and school quality. Our results generally align with theoretical predictions of how schools should react to increased competition. We provide evidence of vertical differentiation and show competition shifts resources towards instructional expenditures. In the long-run, historical school competition has persistent effect on college quality: colleges located in more competitive markets became significantly more selective. While recent studies have documented the unique role played by elite private universities in fostering upward mobility (Chetty et. al 2017), the process by which these colleges become selective is not well understood. Our analysis provides a historical account for the formation of elite universities.

The rest of the paper is organized as follow. Section 2 conducts a literature review and discusses our relative contributions. Section 3 provides a brief history of denominational involvement in higher education. Section 4 explains our contribution to data and the sources of our datasets. Section 5 describes our main empirical specifications and show the results. Section 6 presents the analyses based on the Second Great Awakening, which utilizes quasi-experimental variation in New York State. Section 7 estimates an empirical model of school choice which allows us to recover the preferences for attending college of the students' own denominations. In Section 8, we examine the effect of increased competition on school quality. Finally, Section 9 concludes.

2 Related Literature & Contributions

This paper bridges three distinct strands of literature. First, it contributes to the extensive body of research on the economics of religion and, within that, the link between religion and human capital. This literature has sociological roots dating to at least Max Weber's thesis, which proposed a connection between the Protestant doctrine and work ethic. Recent work has documented the economic success of regions that converted early to Protestantism (Bai and Kung, 2011; Becker and Woessmann, 2009). Becker and Woessmann (2009), Mccleary and Pesina (2012), Cantoni (2013) and Cagé (2015) suggested the incentive to accumulate human capital, or increase literacy, as an explanation.

One particular strand of this inquiry locates the study of Protestantism in the context of the broader role of institutions that affect macroeconomic growth (Acemoglu et al., 2001). For instance, Woodberry (2012) documents the effect of missionary activities on the consolidation of democratic institutions.

A recurring theme along this line of research is that norms associated with specific religion or denomination matter. This is echo'd in evidence from developing settings. Examples include

Geruso and Spears (2017), which explores the Hindu and Muslims cleavage its relevance for sanitation habits, Chaudhary and Rubin (2011) discuss the relationship between reading, writing and religion in colonial India. Other notable recent papers include Kuran (2016), Kuran (2014) and Chaney (2013).[3]

Beyond the consequences of adopting specific religions or denominations, economists have also studied the interactions between denominations or religions more broadly. Interestingly, it was Adam Smith who first referenced the church and competition between religions in *The Wealth of Nations and* and *The Theory of Moral Sentiments*. Smith wrote about religious pluralism and argued that competition benefited the consumers of religion and constrained the extent of rent extraction by religious authorities.[4]

Subsequently, economists have employed the rational choice framework to understand how religious competition matters (Becker, 2003; Bisin and Verdier, 2000; Iannaccone, 2008; Montgomery, 2003; Prummer and Siedlarek, 2014; Putnam, 2000).

Yet the existing literature has been primarily concerned with the effect of religious competition on the religious practices, whether it be religious participation or the shaping of religious customs, less well understood is the ramifications of religious competition on a broader set of outcomes or the formation of institutions. Notable exceptions include Jha (2013) who investigates the historical complementarity between Hindus and Muslims and its effect on the incidence of conflicts.

In this domain, the closest paper to our own is Iyer et al. (2014), which provides evidence for religious competition and cooperation in contemporary India by examining religious and non-religious service provision. To our knowledge, our paper is the first to investigate the relationship between denominational competition and institutions of higher education.

Second, our paper relates to the literature on economic implications of diversity. Studies have typically emphasized the economic costs of diversity. Easterly and Levine (1997) shows that ethnic diversity adversely affects public policies associated with economic growth such as black market, low provision of infrastructure, and low levels of education. Alesina et al. (2003) finds that the provision of public goods such as education, roads, and sewers is inversely related to ethnic fragmentation in US cities.

Our paper differs from this prior literature in several aspects. Thus far, economists have examined on birth, ethnic, and linguistic fractionalization; our paper explores the direct effect

[3]Kuran (2016), Kuran (2014) examines the emergence of *zakat* in Islamic Code and its effect on the development of financial system in the Middle East. Chaney (2013) discusses the relationship between religious authority and political power.

[4]In contrast, Hume contended that state sponsorship of a unique religion was welfare improving.

of religious diversity. And whereas existing literature considers the provision of public goods as an outcome, our paper studies the unique interaction between religious competition and the provision of education services in private markets.

Lastly, because we treat college building as an entry problem, we draw from the industrial organization literature on differentiated products and firm entry in the spirit of Bresnahan and Reiss (1991) and Mazzeo (2002). Iannaccone (1992) introduced the use of industrial organization concepts in the economic analysis of religion. McBride (2008) and Montgomery (2003) use the methods from industrial organization and product differentiation to examine the relationship between pluralism and participation. We build on these insights and additionally conceptualize religious affiliation as providing access to credit or financial networks and markets. We also quantify the preference for religious homophily in the context of 19th century higher education explicitly.

3 Historical Background

The connection between American higher education and religion dates back to the Colonial period. Nearly all nine of the colonial colleges had denominational affiliations: Harvard and Yale were congregational, College of William & Mary and Columbia were Anglican, and Brown was Episcopal, just to name a few.[5]

University services were intertwined with religious functions. Colleges trained and supplied future ministers in addition to providing a formalized curriculum of classical education. More than half of the Harvard graduates became ministers for the sixty years following its founding. Nearly three-forth of the graduates of Yale became ministers for the first twelves years since its founding (Tewksbury, 1932). A steady 17 percent of Princeton graduates became ministers from 1824 to 1854 (Bubolz, 2000).

Ministerial work remained a promising career path up until the period after the Civil War; and university education differed in their specific interpretation of the Bible.[6] Since ministers serve churches that match with their university's denominational affiliation, denominational differences really mattered. The founding of Yale College best illustrates this significance: After Orthodox churches in Connecticut had separated themselves from their sister churches in Massachusetts, Connecticut refused ministers graduated from Harvard. The subsequent found-

[5]Two colleges, Princeton and University of Pennsylvania, were officially non-sectarian but nevertheless had were primarily influenced by Presbyterian and Church of England respectively.

[6]In this time period, this was often the only difference in curriculum as secular education in college was largely a homogeneous good.

ing of Yale, as some evidence suggested (Tewksbury, 1932), was to supply orthodox ministers to Connecticut churches.

Denominational involvement in higher education continued and intensified in the 19th century. Competition among different denominations stimulated the establishment of colleges. Denominational college building frenzy permeated the Old Northwest from 1790 to 1860. In Ohio alone, 43 institutions were founded in this period, including Kenyon (1824), Western Reserve (1826), Oberlin (1833), Denison (1831), and Marietta (1835).

In the absence of alumni network or government funding, denominational colleges depended on the influence of affiliated churches. With an official or verbal sanction, a religious organization legitimized the college founding endeavor and rallied local economic support (Potts, 1971). Local support was crucial, now that colleges served local market: the colleges founded in the early 19th century were in most cases small in size, located in rural towns and recruited students from areas within 50 mile radius(Church and Sedlak, 1997).

Denominational exerted greater control in the Midwest, where state fund or aid for colleges were scare.[7] They formed organizations to make explicit effort in college fund raising. By the order of the Indiana Conference (a Methodist society), all Methodist ministers in Indiana were fund raising agents for Indiana Asbury University, responsible to "solicit funds, procure students, and collect what books the liberality of the public may bestow..." (Findlay, 2000). The Indiana Conference even gave instructions to preachers for more effective solicitation.[8] The Baptist, Presbyterian and Congregationalists societies had analogous organizations to Methodist's.

Besides rallying local support, denominations utilized their national network to secure resources beyond their immediate surroundings. In order to maximize donations from the east coast, denominations planned top-down visiting schedules of their fund raising agents to avoid exhaustion of donor's interests to give. Agents salaries, varied with their performance, were deducted proportionally from the sums collected. According to John Peck, agent of Shurtleff College (Baptist), about 40 to 60 percent of funds gathered reached college offers (Findlay, 2000). The denominational network became a source of finance and allowed institutions to secure resources beyond their immediate surroundings.

The ability to organize fund-raise beyond a local network earned denominational colleges the vantage to survival. This became especially true as formal credit markets became rare and given

[7]In New England, New York and Pennsylvania, institutions received substantial economic aid from state funds throughout the Colonial and post–Revolutionary periods (Burke, 1982).

[8]"the best method of collecting funds, that is, to not to depend on public collections alone, but in their pastoral visits to bring the subject before the members and friends of the church, individually."(Findlay, 2000)

the non-profit disposition of the projects. Ninety percent of Protestant colleges founded before 1860 survived through the Civil War, while only thirty-seven percent of non-denominational colleges and seventy-four percent of state colleges did. (Burke, 1982). In the absence of non-denominational source of funding, college leaders oriented their rhetoric to the religious aspects of their institutions. Denominational affiliation became an existential necessity. The result was the unique American private college system controlled by denominational societies and bodies of laymen.

Colleges were vehicles of religious conversion. Newer denominations built colleges as means of competing for social space with older religions. In the 1800s and 1810s, the three oldest denominations, Congregationalists, Episcopalians and Presbyterians were connected with 70% of all colleges, by the time of the Civil War, the share dropped to one third (Burke, 1982). Church and Sedlak (1997) described how colleges were established as soon as a town or a state was founded at the frontier line. The drive was less of educating priests and more from the concern about losing the allegiance of young men educated by rival denominations. Denominations such as Methodist and Baptists, who were anti-formalists in nature and opposed existing seminaries (Johnson, 2008), affiliated one third of total colleges by 1860.

A contributing factor was an intense period of Protestant religious revival and innovation that took place during the early 19th century. This is commonly referred to as the Second Great Awakening. The movement introduced camp meetings, an outdoor religious gathering that celebrated emotional spirituality. Usually held in rural areas, the camp meetings enrolled new practitioners. Active denominations such as Baptist, Methodist and Presbyterian grew into the dominant positions and saw marked increase in their membership. A number of new denominations such as Seventh-day Adventist, Disciples of Christ, Church of Christ, etc. were born from the midst of the Second Great Awakening. These newer denominations, especially the lower status ones, entered into the competition for higher education.

The tit for tat relationship is probably best illustrated by an example from Ohio:

> Ohio Methodists chafed at Presbyterian control over the two nominally state universities, Ohio and Miami. Their response was to found Ohio Wesleyan College in 1842. Although in the centre of the state, Ohio Wesleyan offered little succor to Methodists in the northeastern section who wished to counter the lure of Oberlin. A wealthy layman, John Baldwin, worked with the North Ohio Conference to found an eponymous institute. Raised to collegiate status in 1854, Baldwin College soon surpassed Oberlin in the wide spectrum of its offerings. A third Methodist college, Mount Union (1858) was entirely the work of laymen. Beginning as a seminary in

1846, it also quickly sprouted additional departments. In governance Mount Union resembled earlier colleges in being firmly associated with a church, but not formally affiliated.

Both the Brethren and the German Reformed Churches were moved to found colleges for the same professed motives as the Methodists. The Brethren had eschewed colleges until they felt pressured into launching Otterbin in 1847. For the German Reformed Church, the long distance to their seminary in Mercersberg, Pennsylvania, was the decisive factor in starting Heidelberg College (1850). The founding of Antioch (1852) might be considered another variation on this theme. Liberal Christians, who opposed denominational distinctions, sought to found a nonsectarian college of a high rank. They consciously sought the best local offer in New York and Ohio before accepting the bid of the small town of Yellow Springs (Geiger, 2000).

4 Data

To relate the extent of religious competition to the establishment of colleges, we construct three new datasets on religious denominations and colleges, assembled from several primary and secondary sources. In this section, we describe the data and introduce the sources.

Religious Competition: We begin by constructing a panel of religious fragmentation at the county level and its variation across time. This is our main explanatory variable. For this purpose, we consult historical information on religious bodies included with the decennial censuses conducted in 1850, 1860, 1870, and 1890. Although the U.S. Census began in 1790, compilation of religious data did not begin until 1850. A balanced sample of 1,925 counties that appear in at least three decades are included in the panel analysis.[9]

From 1850 to 1890, the census enumerators gathered facts concerning the number of churches, their locations, and their seating accommodations. The information are broken down by denominational affiliation. In 1850, the census identified 18 principal denominations. To the best of our ability, we standardize denomination categories across time to account for differences in granularity of definitions. Nevertheless, by 1890 there are 24 denominations, which reflect the religious growth and innovations that characterized the period. Table 1 reports the average and maximum share of denominational accommodation at county level for each denomination.

The Baptist and the Methodist dominated throughout the decades we focus on, representing over half of the religious share. They were followed by three distant competitors: Presbyterian,

[9]We acknowledge that changes in county boundaries would bring inconsistent units of observation. Unfortunately there is no perfect way to address this.

Episcopalian, and Congregationalists. Even though these three enjoyed the prime status of state supported religions in the colonial time, their influence waned as their authority passed. Lutherans, Catholics thrived as massive immigrants flew into the country. However low in national popularity, the maximum number shows the possibility for almost any denomination to exercise control locally. The dynamics over the decades and space created considerable variation in religious composition across counties and over time, making our panel analysis feasible.

To measure religious competition, we use church seating accommodation as a proxy for market share. We compute denominational competition as 1 minus a Hirschman-Hirfindahl index of the shares of each denomination's accommodations in total accommodations. Intuitively this is a measure of market concentration and the value is greater in counties where the religious market isn't very concentrated. Specifically the denominational competition at county c is:

$$DenomFraction_c = 1 - \sum_i s_{ic}^2 \tag{1}$$

where s_{ic} is the share of denomination i's accommodation as fraction of total church accommodation in county c. Figures 1 - 4 shows the spatial distribution of denominational competition from 1850 to 1890: it was high in the Northeast and Midwest, but also in the Gold Rushed California, and in the frontier states like Colorado. Beyond regional disparities, there were also sizable differences across counties in each states, allowing identification from within-state variation.[10]

College Information and Location: Compiling from a number of sources, we build an original dataset of 19th century US colleges. In 1850, the U.S Census recorded the number of colleges in each county. For 1860, we rely on enumerations of colleges by two historians. In his doctoral thesis *The Founding of American Colleges and Universities*, Donald Tewksbury catalogued antebellum colleges from state legislatures and charters.[11] The list of *permanent* colleges was published in the thesis, but the list of failed colleges was lost. Permanent here means surviving to 1920, thereby the Tewksbury colleges are a subset of all operational colleges in 1860.[12]

To complete the list of colleges standing in 1860, we transcribed additional data from *The Founding of American Colleges and Universities* by Colin Burke. His method of collecting data

[10]We also see there is substantial spatial sorting and clustering along denominational lines. Even for small denominations, in term of national representation, there counties where they are the majority.

[11]The building of even private colleges required state approval.

[12]In his finding, the average Antebellum college mortality rate was as high as 81 percent for the sixteen states of the Union.

contrasts with that of Tewksbury. He utilized city registries, phone books and any proof of existence he could find. He then investigated whether a college actually taught collegiate level of courses and combine the ones that did into a list. Figure 5 shows a typical entry in Burke's.

The entry contains detailed information of each college. Both authors emphasized the denominational affiliations of colleges and recorded them when existed. Based on the union of colleges identified in the two sources, we construct and geocoded an entirely new and comprehensive dataset of colleges along with information on their location, history, and denominational status. The data is then tabulated to provide county level counts of colleges by 1860.

The *Report of commissioners of education* provided rather detailed information on colleges since 1870s on an annual basis. In these reports, the commissioner conducted census of institutions of higher education on a voluntary basis through surveys. Extensive information were collected including: names, location, denomination affiliation, founding date, endowment, land value, library volumes, enrollment, expenses, tuition and board. All the information was self reported from each college. To address and mitigate problems associated with non response, we proximate missing data by filling in information provided in adjacent years. Altogether, we are able to assemble a rather complete set of colleges for 1870 and 1890.

Figure 6 plots the total number of colleges in our compiled dataset from 1850 to 1890. We were able to identity denominational colleges from 1860 to 1880. In the category of nondenominational colleges, we divide it into public and private. Between 1860 and 1890, the number of denominational colleges grew by more than 35%. The growth rate, from 1870 onward, is offset by the concurrent the secularization of existing denominational colleges. The new colleges built in this time period were overwhelmingly denominational.

Students Micro-Data & College Choices: To our knowledge, little is known regarding the demand side of colleges in the 19th century. Because of the lack of linked administrative data dating from this period, there's a scarcity of information on who went to colleges, the choice of colleges, and factors determining those choices. We overcome this challenge and introduce a unique source of matched student college data to the literature.

We acquire a dataset on Antebellum college student dataset from Colin Burke. He surveyed large array of related materials held in the Library of Congress, the library of almost all universities and colleges, alumni registers, yearbooks, and other archival records of the colleges. From the text of the individual biographies included in those documents, Burke extracted rich demographic information about the student and their educational path.

The original dataset contains about 12,000 students, with their names, hometown zip code, college destination, and known denomination affiliation. Occupation, father's occupation as

well as extra curriculum were recorded whenever available. And after omitting observations with missing data, we retain around 3,000 students with full set of covariates.

Summary statistics of all students and colleges appearing in the sample are provided in Table 2.

The first observation of interest is that students on average traveled a relative short distance, of 200 kilometers, from hometown to attend college. This reinforce the notion that colleges served local and isolated markets. Thereby it makes sense to consider a county the relevant market definition and the within county variation in religious competition the pertinent demand factor. This substantiates the county level analysis in the following section.

Second, we see the intuition charged by colleges was relatively low. Higher education in the 19th century, even amongst the elite institutions, is best characterized as a buyers market. Pool of applicants were small and admission was not competitive. Colleges like Columbia and Harvard regularly struggled to fill their fall class and put out advertisements to do so steadily through August or September, just days away from the semester's start.[13] The strong competitive forces exerted downward pressure on tuition.

Historical Controls: We gather economic and demographic county characteristics covering the period, 1850-1890, from the US Census. Haines (2010) provides decadal, county-level, data on manufacturing, agricultural production as well as migration and demographic information for each county, from the Census of Population, the Census of Agriculture, and the Census of Manufactures. Transportation data are from Atack (2013), which are linked with shape-files of United States county boundaries to account for the spread of railroads and canals.

To control for internal migration and immigration, the share of foreign born and out of state population are calculated from the 1% sample micro-data released by IPUMS.

4.1 Denominations & Supply of Colleges

Before turning to the main empirical analysis, we document the patterns of denominational college building central to our research design. Using the cross section of colleges in 1860, we analyze the determinants of denominational affiliation. In particular, we examine whether the affiliation decisions were *strategic*, i.e. if they responded to local demand factors.

Table 3 shows that the denominational composition within the county plays a significant role in driving observed supply. We provide results for the six denominations that were the largest total accommodations. A unit of observation is a college in 1860. The table presents OLS regression of a dummy for whether college chooses affiliation to be certain denomination (Baptist

[13] See: https://thechoice.blogs.nytimes.com/2011/03/31/remembering-when-college-was-a-buyers-bazaar/

in column 1, Methodist in column 2, and Presbyterian in column 3, etc.) on market shares of those denominations while controlling for geographic and socio-economic controls. Given the construction of the dependent variable, the coefficient can be interpreted the marginal effect of a denomination's market share on the likelihood of a college's affiliating with the denomination, conditional on the existence of a college.

We find that colleges aligned themselves with the dominant denomination in a given county. For instance, the larger the Baptist share of denomination accommodation in a county, the greater the possibility of a Baptist college. Methodist and Presbyterian denominations exhibit the same pattern. The estimated coefficient for the Episcopalian share on college affiliation is the largest among the six, with an elasticity of over one. This implies a 10 percentage point increase in the Episcopalian share increases the likelihood of a Episcopalian affiliation by 16 percentage point.

The results can be rationalized by a standard model of entry and competition where the profitability and survival of a college depends on whether enough students can be recruited locally and whether enough financial support can be raised. Under the assumption that a student derives higher utility from enrolling in a college that matches with his own denomination, and that fundraising depends heavily on local network, a college's likelihood to adopt a denomination should increase with its share amongst the local population. This provides preliminary evidence on the salience of affiliation choices.

5 County Level Empirical Analysis

This section presents the empirical exploration of the relationship between religious fragmentation and college establishments through OLS estimation. A discussion of endogeneity concerns and corresponding instrumental variable solution is presented in the next section.

5.1 Cross Section

The cross section estimating equation is:

$$log(\#\ of\ colleges_c) = \beta DenomFraction_c + \sum^{j} \gamma_j Accommodation_{j,c} + \lambda_s + \sigma X_c + \epsilon_c, \quad (2)$$

where the outcome is the 1 + total number of college in county c in logarithm, $Accommodation_{j,c}$ defines the sitting capacity of a denomination j in county c, $DenomFraction$ is our constructed Herfindahl Index of denominational competition that equals $1 - \sum_j Accommodation_j^2$, λ_s is a state fixed effect, X_c is a vector of control variables, and ϵ_c is an error term.

The coefficient of interest, β, captures the relationship between denominational rivalry and college counts. A positive β provides evidence that competition stimulated denominations to establish more institutions.

We consider different specifications that successively expand the set of controls to include state fixed effects, denominational fixed effects, geological controls, and social economic controls. Definitions of these social-economic controls as well as their sources are given in the appendix.

We control for state fixed effect and cluster standard error at the state level in order to control for unobservables that affects settlement pattern for both denominations and colleges that are geographically fixed. This is intended to capture sharp regional differences in habitability, availability of public funds, social structure (slavery practice) and some degree of cultural attitude. Column 1 in Table 4 shows this baseline estimate.

We add denomination controls γ_j to account for the dominance of specific denominations. By construction, our religious Herfindahl index is based on the squared values of denominational shares. Some denominations (Baptist, Methodist, Presbyterian, Congregation, Episcopal) represented large shares of religious market, and these shares may be strongly correlated with religious competition. Since denominations vary in their attitudes towards higher education, we address this by controlling denomination shares and including dummies for each of the largest five denomination that takes a value of one when the denomination has the largest share in the county's religious market. By including these fixed effects, our coefficient on religious fragmentation is only identified from variation in the composition of religious bodies as opposed to the presence of any single denomination. Column 2 in Table 4 shows that results are robust to controlling for dominance in denominations in different ways.

Geological controls include county area, agricultural productivity, the (Euclidean) distance to the coast, distances to great lakes, and distances to major metropolitan cities. Including geological controls is important as they affect the likelihood of settlement for both denominations and candidate college. Distance to coast, waterway and metropolitan cities proxies access to market, which could potentially increase religious diversity as access to market is known to increase trade, population and income. Column 3 of Table 4 show estimates after controlling for Geological conditions.

We consider three subcategories of social economic controls: demographic characteristics, industrialization, and transportation infrastructure. The first category consists of the shares of population corresponding to people in urban area, out of state migrants, foreigners, male, white, and aged 5 to 18. These population controls account for different trends in migration and population growth which may have been correlated with the propensity to build colleges.

We measure industrialization level by using manufacturing employment, output, and investment. Transportation controls include a dummy for railroad access, miles of railroads, and a dummy for access to steam-boat navigated rivers. Column 4 in Table 4 shows results in this specification. A large literature has confirmed that railroads and canal network have significant implication on banking, urbanization and industrialization. Many anecdotal evidences also suggested railroad connection as a key factor of consideration when college founders decided on candidate locations. Thus, the robustness of the results to controlling for the transportation variables is quite reassuring.

5.2 Panel Specification

To conclude this empirical exploration, we examine the hypothesis in the panel framework. The analysis follows the specifications in the cross section with a few modifications. We introduce time variation and exploit the extent that religious fragmentation varies at different points in time to investigate the effect of changes in religious diversity on changes in number of colleges. In contrast to the previous analysis, here our estimates are identified not by regional variation in the cross-section, but variation across time in a given location.

The estimating equation is:

$$log(\#\ of\ colleges_{ct}) = \beta DenomFraction_{c,t} + \sum^{j} \gamma_j Accommodation_{j,c,t} + \lambda_c + \sigma X_{c,t} + \epsilon_{c,t} \quad (3)$$

Formally, we replace state fixed effects with county fixed effects λ_c. Our panel dataset contains 7,276 observations. Since the U.S incorporated 14 new states from 1850 to 1890, we choose to perform the analysis in an unbalanced panel to best capture this dynamic period. A county is included in the panel if it was observed in at least three out of four decades.

Column 1 in table 5 shows results of the baseline panel regression. Standard error are clustered at county level to account for spatial autocorrelations across time. The inclusion of county fixed effect allows us to flexibly address any unobserved time invariant county characteristic which may differentially affect growth or attract settlement.

Column 2 shows estimates with additional denomination controls. Since most geological controls are fixed over time, we exclude all but agricultural productivity from the estimation. Results controlling for farm productivity is reported in Column 3.

Follows the same line of reasoning, access to steam-boat navigated rivers is excluded. We expect little changes in canal building from 1850 to 1890, as most canals were completed by early nineteenth century. Besides that, all social economic controls as well as railroad

infrastructure controls were kept and estimates from including social economic controls are shown in Columns 4 & 5. Lastly, by including the decadal fixed-effects, the treatment effect of religious fragmentation is only identified from variation within a county within the decade.

In conclusion, the OLS analysis indicates a large, positive and robust relationship between denomination fractionalization and growth of operating colleges exist both *between* counties and *within* counties across time. A back-of-the-envelope calculation using the estimates from panel data reveals that if the U.S. was dominated by one religion, there would have been approximately 20 percent fewer colleges by 1890. Because the estimate is identified off within sample variation from 1850 to 1890, to the extent that college building in 1850 already reflected prior denominational fragmentation, this will understates the overall historic impact of religious competition.

5.3 Discussion & Interpretation

In this section, we discuss the implications of the OLS estimates and address potential threats to identification.

Several factors could potentially contaminate the OLS estimates. In the preceding section, we try to address competing explanations directly by explicitly accounting for them. The controls excluded differences in population, demographics, manufacturing, transportation networks, land productivity, migration, religious norms, etc as confounding stories or explanations. And because of the panel analysis, which relates changes in colleges to changes in religious diversity, we are able to flexibly control for time-invariant locational fundamentals. Nevertheless, we acknowledge these precautions are not exhaustive and important concerns remain.

On the one hand, a college served as a booster to the local economy by bringing migrants and commercial opportunities. It is natural to suspect that a college would attract a diverse body of migrants, which gives rise to a bias due to reverse causality. This is a plausible concern, as historical narratives often quote that colleges marked the arrival of civilization and signaled uplifted moral standard. Given this, the content of the institution likely mattered less than the symbol of its physical presence. Some would even suggest to treat colleges in the same spirit as canals, railroads and steam-boat rivers (Goetzmann, 2009).

We speak to this issue by testing whether having colleges in 1860 lead to an more religious fragmented county in 1890. Table 7 presents the result in full sample as well as subsamples that represent different regions. There was no evidence that colleges contributed to religious diversity. The subsequent section on college demand will provide insight on this outcome.

Since an average student preferred to attend college that matched with his own denomination, a college might bring people but people who were homogeneous in belief.

On the other hand, the positive association between religious competition and college founding could be driven by omitted variables that induced higher fragmentation and also demand for higher education. For example, religious diversity may reflect cultural tolerance and a high propensity to adopt new ideas, which would result in a higher demand for colleges.

To alleviate concerns about this selection, we test whether denomination fragmentation correlates with establishments of state or land-grant universities. The experience of founding public universities assimilate to that of denominational colleges, in the sense that a community's willingness to support it mattered significantly.[14] If it were true that demand of higher education was the unobserved driving force, then applying the OLS regression on land-grant colleges would yield positive result.

To address this formally, we collect location data on existing state, land-grant universities and additional land-grant colleges that were designated to be built in 1890. We run log-linear regressions on these public institutions using 1890 religious and social economic data.

Table 6 shows results of this placebo test. Columns 1 & 2 report an null coefficient on denomination fragmentation, whose size is further taken away when the existence of state colleges is controlled for. Column 4 narrates the same pattern when the dependent variable is the combined total of state and land-grant colleges.

We extend our test further to include a regression on the number of public high schools in 1850. The financial support for secondary education differ from that of universities. Since funds for public high school were collected from real estate taxes regardless whether a family had children to sent to school, a key determinant for public school demand is homogenous population (Goldin and Katz, 2009). The homogeneity spans dimensions on income, parental education level and, on what matters to us, religion. Therefore, we expect a negative relationship between the religious diversity index and density of public high schools. Column 3 in Table 6 is consistent with this idea.

While the estimates from public high school regression validate the identifying assumption, it introduces another endogenous concern. If social capital is important in the story of college founding and social capital is negatively associated with religious diversity, then the OLS estimation risks a downward bias. And this can attenuate the results.

[14] In his dissertation, Andrews (2017) described the petition process of land-grant colleges

6 The Second Great Awakening

What drives the underlying variation in denominational fragmentation? Fundamentally, whether we can interpret our estimates as causal or merely correlational depends on the answer. An unfortunate shortcoming of our research design is the lack of transparency with regard to what explains the variation in our explanatory variable. Religious diversity is certainly an endogenous outcome of a complex set of historical process. Even though we control for correlates to the best of our abilities, the issue of joint determination in our research question is of particular concern. Remaining unobserved characteristics are serious threats as they might bias, or in the worst case invalidate, the hypothesized relationship between religious diversity and higher education.[15] To resolve these latent ambiguities, we focus on one source of variation in religious fractionalization: religious revivals attributed to the Second Great Awakening, and isolate plausibly exogenous variation there within.

As mentioned in the historical background, early nineteenth century witnessed the spread of religious revivalism commonly referred to as the Second Great Awakening. The Second Great Awakening was a national phenomenon that began in the late eighteenth century and continued through the first third of the nineteenth. Americans in virtually every town or county experienced religious revivals and organized evangelical voluntary society.

The United States, in contrast to most established countries at the time, had no governmentally established church monopoly. Hence, religious proselytizing began in the colonial period and accelerated in the early national period, rising to a fevered pitch during the Second Great Awakening. During this period, membership for leading evangelical denominations grew twice as rapidly as population growth (Finke and Stark, 1992) while newer denominations also found opportunities to enter the market.

The Second Great Awakening was characterized by numerous revival meetings, whether they were the orderly preaching organized by the Formalists, or the ecstatic camp meetings set up by the Anti-Formalists.[16] The revival events likely reshuffled the rank of popularity of denominations, as well as the religious composition of various locations. These camp meetings led to extraordinary numbers of people to convert through an enthusiastic style of preaching and audience participation.

[15] Religious diversity could be the result of migration and concentration of people of different cultures and abilities in regions characterized by high economic potential. We control for migration explicitly but counties with pronounced diversity might also be characterized by a more liberal socioeconomic environment, open to new cultures and ideas.

[16] Formalists composed of elite Presbyterian, Reformed Dutch, and Congregationalists; Representatives of Anti-Formalists were Methodists and Baptists.

The revival events reflected Romanticism ideals in their enthusiasm, emotion, and an appeal to the supernatural. Religious leaders in this period rejected the skeptical rationalism of the Enlightenment and the movement had anti-intellectual elements; the contents of revivals had little intellectual value. Given the motivation for revival and camp meetings was intrinsically religious and spiritual, we interpret them as natural experiments on the character of local religious composition. In this section, we utilizes several datasets from New York to study the "reduced-form" effects of revivals on college establishments.[17]

6.1 Data & Identification Strategy

Our data on revivalism comes from Hammond (2007), who collected information on religious revivals in New York between 1825 to 1835 from religious newspapers. For each instance, he recorded the location, year, and number of people affected.[18]

Hence our empirical study is confined to a decade long period in one state. However, the choice of the period and locations is not arbitrary. New York, and especially upstate New York, was a particularly important region during the Second Great Awakening. Western New York was christened the "Burned-Over District" by nineteenth-century contemporaries because of the frequent occurrence of spiritual revivals. The religious enthusiasm in this part of the country "burnt" hotter than many comparable regions.

Furthermore, the period covers the influential Rochester Revival (9/3/1830-3/3/1831), which is considered a significant point in narrative history of the Second Great Awakening. This particular revival, organized by Pastor Charles Finney, was noted for introducing several innovations into revival practices and inspiring zeal in nearby towns in the subsequent year, when revival activities reached its arguable zenith. Figures **??** & **??** shows the yearly trend in revivalism activity by total number of events and total number of converts, respectively. In accordance with this account, the plots show a significant spike in activities during 1831.

We organize our analysis at the township level, which is the finest geographic information provided, and aggregate the sum of revival events and affected population by township. We acquired digitized demographic data from Rogers (2010). The original data was taken from the U.S census of 1820 and the census towns were merged with revival locations referenced by

[17] The relationship we study is "reduced form" because we omit the intermediary channel of denominational fragmentation for the sake of empirical parsimony.

[18] His newspaper sources are: Home Missionary and American Pastor's Journal (New York, 1828-1831), Methodist Magazine (New York, 1825-1828), New York Observer (1825-1835), New York Evangelist (1832-1834), Rochester Observer (1827-1832), Western Recorder (Utica, 1825-1826), Utica Christian Repository (1833), The Christian Advocate (New York, 1826-1835), The Evangelical Magazine and Gospel Advocate (Utica, 1830-1831), and The Visitant (Utica, 1825-1827). Unfortunately, denomination was not specified in each revival entry.

19th-century gazetteers (French, 1860; Gordon, 1836; Spafford, 1824). A total of 521 townships was made available for study, with variables in population size, share of population corresponds to manufacturing, agriculture and commerce, and total area. We geocoded the towns to add transportation and geological controls. Distances to the canals and waterways at 1825 are computed using shape-files from Atack (2013). Geological controls include distance to major cities and mean elevation.

Table 8 presents descriptive statistics of NY towns. Columns 1&2 report the mean and standard deviation of variables for towns with at least 1 revival meeting(revival towns); Columns 3& 4 report those for all NY towns.[19] Revival towns are slightly more populous and industrialized than average New York towns.

Identification Strategy:

While the expressed purpose of revival events had little connection to higher education by themselves, they obviously did not occur exogenously and their locations were not randomly chosen. To remove lingering concerns over the endogenous selection of religious activities and eliminate potential biases, we adopt an instrumental variable framework.

We focus on incidents of revivals in 1831 specifically, as that year constituted a major shock in the flow of revival actives as shown in Figures **??** & **??**. Two facts motivate our choice of instruments. First, a key feature of revivals is that they were predominantly outdoor events and therefore subjugated to the constraints of weather. We take advantage of this intuition and utilize precipitation in 1831 as an instrument for revival activity. The relevance assumption is that in places that experienced more rainfall in 1831, fewer opportunities to organize revivals existed and for revivals that did place turnout would presumably be lower. And the exclusion restriction is that transitory rainfall shocks in 1831 had no direct effect on higher education investment in later years.

Second, we leverage the fact that Charles Finney's Rochester Revival inspired followup revivals in nearby towns. We hypothesize that the strength of its influence would be most strongly felt in areas immediately bordering Rochester and expanding from there onward. Thereby, we construct the Euclidean (straight line) distance from each town in our sample to Rochester and utilize it as an additional instrument.

Ultimately, we combine these two sources of exogenous variation in an instrumental variable model to estimate the causal effect of religious revivals on college establishment in the subsequent decades. Coincidentally, rainfall pattern in 1831 correlates largely with distances to Rochester: towns with the highest volume of precipitation are most distant to Rochester.

[19] New York City and Albany were excluded

To increase identification power, we regress 1831 rainfall on distance to Rochester and utilize both 1831 rainfall and the residual distance as instruments. We estimate the following system of equations by 2SLS.

$$Revivalism_{j,1831} = \sigma_1 \cdot Precipitation_{j,1831} + \sigma_2 \cdot Distance\ to\ Rochester_j + \pi X_{j,1830} + \upsilon_j \quad (4)$$

$$log(\#\ of\ colleges_{j,1860}) = \beta \cdot Revivalism_{j,1831} + \gamma X_{j,1830} + \epsilon_j, \quad (5)$$

where the endogenous treatment, $Revivalism_{j,1831}$ is the incidences of revivalism activities in town j in year 1831. The outcome $log(\#\ of\ colleges_{j,1860})$ is the number of colleges constructed in town j by 1860 and β is the coefficient of interest. The excluded instruments in the first stage are $Precipitation_{j,1831}$, the frequency of rainfall in 1831 in town j, and $Distance\ to\ Rochester_j$, the residual Euclidean distance between Rochester and town j from a linear regression of $Precipitation_{j,1831}$. Finally, X_j is a vector of town level socio-demographic and geological characteristics as well as the history of revival activities in the town prior to 1831.

6.2 Empirical Results

Before turning to the 2SLS results, we examine the first stage relationship in our township sample. Column 1 & 2 of Table 10 shows the the coefficient on the instruments in predicting the number of revivals and number of converts in 1831. The relevance assumption is clearly satisfied as towns with more rainfall experienced fewer revivals and towns closer to Rochester experienced more. However, the interpretation of β as the causal effect of revival activities in the second stage also requires that the exclusion restriction holds. The credibility of our research design hinges on the assumption that days of rainfall in 1831 and distance to Rochester did not affect college construction directly or indirectly for reasons other than revival activities in that year.

While the exclusion restriction is fundamentally untestable, we address a few threats to identification. One potential violation of the exclusion restriction is the direct effect of rainfall on economic development. Areas that receive prolonged period of high precipitation will likely differ economically from areas that do not. To the degree that annual rainfall figures are serially correlated, a concern is that 1831 rainfall is not transitory in nature but reflect average annual rainfall over long period of time. To test this, in Column 3 we add in the 1830 & 1832 precipitation data as an additional explanatory variable. Reassuringly, we see it is the only the 1831 rainfall that explains the variation in revival intensity. This indicates it is not the

average rainfall which drove religious activities but the idiosyncratic rainfall deviations during 1831 that mattered. And this identifying variation is likely uncorrelated to long run economic outcomes.

In Column 4, we assess if residual distance to Albany, the New York State capital, determined the strength of revival activities in 1831 and we see the relationship is insignificant. This bolsters our claim that the relevance of distance to Rochester is driven only by the actions of Pastor Finney as opposed to some broader effect of market access, in which case Albany would be equally consequent. This reinforces the exogeneity of our instruments.

Now we turn to quantify the marginal effect of revivals on college establishments in the second stage. Columns 1 & 2 of Table 11 show that IV estimates. The results are robust to controlling for geological and pre-existing demographic conditions. Since future revival events likely correlate with 1831 revivals in location, the aggregate impact of revivals on college establishment will be larger. We substantiate our results by testing effect on colleges built before 1825 as a falsification experiment. This is to address the concern that if there are other factors correlate with proximity to Rochester affecting the probability to build a college, our exclusion restriction is threatened. We're reassured to find no significant effect on previous colleges by the result reported in Column 5.

The IV estimates are larger in size and more significant than the OLS estimates, which are reported in Columns 3 & 4 of Table 11. The discrepancy between IV and OLS estimates point to the fact that revivals could be correlated with omitted variables that are negatively associated with investment in higher education, resulting in downward bias in the OLS estimates. This suggests that if anything there is negative bias in the selection of revival with respect to educational investment. This accords with the narrative evidence which indicate that revivals were spiritually motivated and largely took place in rural communities that possibly placed less emphasis in higher education. The bias could occur, for instance, if a college was established thanks to intense revivalism in adjacent towns. Both the college town and adjacent revival towns would share the revival enthusiasm, and yet one is more suitable to build a college campus and the others with natural settings more suitable for revivals. The difference between OLS and IV estimates could also be due to attenuation bias induced by the measurement error in revival data. A newspaper source was usually specific to a location and denominational affiliation, so the geographic coverage of revival events could be incomplete and biased. The instruments correct both sources of bias. Although the magnitudes are large compare to OLS estimates, they are reasonable in an absolute sense.

The analysis in this section offers two distinct advantages. First, it is organized at the township level which is a much more granular geographic definition. Given that the forces we are emphasizing and the story we are telling is fundamentally a local one, the township is a more pertinent and appropriate unit of observation. Second, the empirical design make it readily apparent what the treatment is. This allows us to assess the plausibility of the identification strategy more easily. Whereas it was difficult to consider the excludability restriction in the prior section because of the confluence of factors that contribute to religious fragmentation, revivals have specific historically motivated origins that are easy to ascertain. Altogether, this allows us to make progress toward actual causal estimates of the effect of religious diversity on higher education.

7 Student Preferences & School Choice

The United States is unique in its absence of state sponsored religions and, consequently, the proliferation of religious denominations and factions in the 19th century. The results in the preceding sections indicate that, incidentally, this historical fact had unintended and unplanned consequences for the growth of higher education.

In this section, we attempt to provide some evidence on the potential mechanisms which underlie these findings. We consider the denominational affiliation of a college as a product characteristic and a *strategic* choice by the school. This dimension of horizontal differentiation allows universities to cater to the preferences of the consumers exactly. And to the degree that the underlying taste in the population is heterogenous, this will be reflected in the increased provision of varieties within a decentralized market. Colleges can avoid competing in prices by maximally differentiating themselves along the religious spectrum. Importantly, the gains to differentiation from these standard Hotelling channels allows for a greater number of entrants to be sustained in equilibrium. This intuition can be formalized in models akin to Seim (2001) and Gentzkow et al. (2011).

The fundamental assumptions central to this demand driven explanation are: 1) colleges serve local markets 2) consumers exhibit preference for schools with denominational affiliation that match their own, and regard colleges with the same denomination as more substitutable than colleges with different denomination. We assess the strength of these assumptions and the preference for religious homophily by estimating an empirical model of school choice. Did students actually prefer to go to college affiliated with the same denomination as their own?

For this purpose, we utilize a sample of linked 19th century student-college data where the college choice and student demographics are jointly observed. The sources of this data is

described in some detail in the Data section. It contains rich demographic information about the students and we merge in college characteristics from the Reports to the Commissioner of Education. Our sample consists of students who attended undergraduate institutions and whose hometown was observed. This sample restriction provides the minimum information required to calculate distance students traveled to attend their respective college of choice.

We measure students' preference to attend schools with same denomination by estimating discrete choice demand model that uses the college choices, along with data on each student, to estimate preferences for school characteristics and how they vary in the population. The empirical model is the familiar conditional logit framework for discrete choice described in McFadden (1974), applied to a setting in which students choose the college to attend.

The framework is based on expected utility framework where students derive utility from attending colleges. Let U_{ij} be the expected utility of individual i from attending school j. Then we suppose that student i chooses school j that maximizes his or her utility over all possible schools in the choice set:

$$U_{ij} > U_{ik}, \ \forall \ k \ \in \ \{1, ..., J\} \ and \ k \neq j \tag{6}$$

Where U_{ij} represents utility over a vector. We assume that it is a linear function of observed student and school characteristics, X_{ij}, plus an unobserved component, ϵ , that reflects unobserved idiosyncratic preference of student i for school j:

$$U_{ij} = X_{ij}\beta + \epsilon_{ij} \tag{7}$$

We assume that the unobservable component, ϵ_{ij} is distributed i.i.d. Extreme Value Type I, which yields the usual logit form for the conditional choice probabilities and allow us to recover the utility parameters.

Several variables comprise the baseline X_{ij} vector, starting with an indicator variable for if the denomination of the college matches the denomination of the student. We are particularly interested in the coefficient on this variable. It reflects the marginal likelihood of attending a school attributed to the religious conformity between the student and school. X_{ij} also contains other alternative specific characteristics such as the distance from home to the college (in km), tuition charged by the school (quoted in per semester figures), size of the school in term of total enrollment, and the quality of school as measured by size of faculty, the volume of books in the library, as well as the founding date of the college.

We also estimate alternative specifications where we add additional fixed effects. First, we include county of origin fixed effects. This is a flexible way of controlling for unobservable characteristics related to the student that correlated with place of origin. Second, we also include county of college fixed effects, this is a proxy for the unobserved labor market conditions or career opportunities at the location of college, which could have driven admission. Lastly, we use college fixed effects instead of alternative-specific characteristics such as tuition or school quality as a robust control for .

We conduct our analysis on two samples of students: 1) the sub-sample of students whose denominational background was recorded; 2) the full sample of students including those with missing denominational information. The first sample is a selected sample of the entire population. From the summary statistics, this group of students differ on several key observable characteristics from the overall population of students. Thus, the estimates based on this group is likely not representative of the true preference parameters and more akin to an upper bound. With the full sample of students, we conservatively assume every student whose denomination information was not recorded is non-religious and did not attend a college of their own denomination. This bias the results against ourselves and provide conservative lower bound on the religious preference of the overall student population.

The baseline, and alternative, estimates are reported in Table 12. In the first column, we include only an indicator variable for religious homophily and the distance between student home town and college in kilometer. In the subsequent specifications, we include successively more characteristics and fixed effects discussed above.

The coefficient on the indicator variable for the college being affiliated with the same denomination as the student is consistently positive and robust across different specifications. A positive sign indicates that a college having the same denominational affiliation as the student is associated with an increase in the probability of the student choosing the college, conditional on other attributes. All other coefficients retain the expected signs: students are more likely to choose colleges that are closer to their home, charge less tuition, established earlier, have more students, with more instructors, and more volume of books in the library.

Coefficient estimates in these models do not have a direct interpretation in terms of magnitude, but the relative size of the coefficients is informative. We derive the marginal willingness to pay (WTP) as a ratio of the coefficient of the non-price attribute of interest to the coefficient of marginal price or tuition.[20]

[20]This is standard practice in these models, as the ratio is comparable to the marginal rate of substitution (MRS).

For example, the estimated marginal WTP for attending a college with the same denomination as the student is calculated as: $-\beta_r/\beta_p$, where β_r and β_p are the coefficients on the religion and tuition terms respectively. Intuitively, this WTP represents the marginal dollar value that a student is willing to spend per semester to attend a college that is affiliated with the denomination identical to his own.

Similarly, we derive the marginal willingness to travel (WTT) as the ratio of the non-distance variable to the coefficient on distance from home to school. This is interpreted as the additional kilometers a student is willing to travel for the the corresponding feature.

For the sample, we find in terms of both WTP and WTT, school denomination was a valued amenity. Depending on the specification, students on average have a marginal WTP of \$512-\$875 per semester for attending a college with identical affiliation. Analogously, students were willing to travel up to 248 kilometers further to attend a college with matching denomination. These figures are large given that the average payment and distance traveled to colleges were only \$54 and 210 kilometers respectively.

Evidently, students derived high utility from attending colleges with affiliation matching their personal denominations. Denominational affiliation was a significant consideration in determining college choice. This implies that there was substantial returns or gains on part of colleges to establish denominational specific colleges to cater to each denomination, and consequently the entry of colleges would be increasing in the diversity of denominations. The strong revealed preferences for same denominational schools suggest high returns to religious differentiation on the part of the college. The effect of religious competition on the market structure of higher education is at least partially explained by the availability of religion as a dimension of differentiated product.

8 Short-run & Long-run Effects on Quality

Thus far, the analysis has focused on how religious competition added more denominational institutions. A remaining question is whether this was a positive development. Even though the quantity of institutions was high, their capacity and quality were questionable. Colleges in the 19th century U.S enrolled small number of students and offered a very limited curriculum. In 1880, average students per private college was fewer than 200 and average faculty per private college was fewer than 10. These colleges more resembled preparatory high schools than the private research universities of today. Was it efficient to have many geographically disperse colleges?

Theoretically, competition among schools can have many positive outcomes, including lower tuition and high quality. However, if institutions are subject to increasing returns then inefficiently high entry could have prevented them from achieving the optimal economy of scale. Which forces prevailed ultimately becomes an empirical question.

We address it by tracking the short run and long run development of denominational colleges. First, we document, in figure 10, the growth of college capacity during the period from 1880s to 1910s, during the onset of the Second Industrial Revolution. The once small colleges expanded their sizes and curriculum to absorb the rising demand for higher education stimulated by the Second Industrial Revolution. Then, we show colleges located in more religious fragmented counties had higher growth rate in capacity and exhibited higher quality in teaching and research, which is consistent with mechanisms such as vertical differentiation or agglomeration effects.

To do so, we construct short run measures of growth and quality using data from the Annual Report to the Commissioner of Education. We take reports every 5 years from 1870 to 1910 plus 1914, the last report in decade 1910. From each report, we transcribed the name and location of each institution, its total students, total faculty, total endowment, enrollment of graduate students and majors offered[21]. About half of colleges in this period experienced some changes in names due to expansion, merger or relocation etc. For instance, Yale College switched into Yale University, Alabama Baptist Colored University renamed as Selma University, and Iowa College became Grinnell College, just to name a few. We investigated each case of ambiguous name change by consulting several secondary sources[22]. Eventually, we were able to compile a panel of colleges spanning from 1870 to 1914. The growth variables of interests are growth rate of students and faculty. The quality measures are indicator of having graduate school and number of majors offered by 1910s.

In the next step, we link our historical college panel to 2018 college ranking by U.S News. We construct two long run quality measures: the indicator of being a research university and the ranking from U.S News. A research university is identified as "National University" according to U.S News.

Table 13 shows differentiated short run outcomes for colleges in more religious fragmented places, measured by DenomFraction. Columns 1&2 estimate that if DenomFraction in 1870 increases by 0.1, the underlying college will have 33 percent more students by 1890 and 30 percent more faculty for the same undergraduate enrollment. Columns 4&5 indicates a higher level of research endeavor thanks to denomination competition. A 0.1 increase in DenomFrac-

[21]Number of graduate students and majors offered were reported in 1985, 1990 and 1914 only

[22]Sources include Burke 1982, Wikipedia and Phonydiploma.com

tion in 1870 raises the chances of having graduate school by 6 percent by 1910; it also induced a wider range of majors being offered by 1910.

More faculty and indicator of graduate program in 1910 predict higher probabilities of being a research university today. Table 14 presents robust effects from both variables. Having established graduate school in 1910 increases the chance of becoming research university by 10 percent. Having hired 20 more faculty increases the chance by 5 percent. This casual evidence suggests that competition may have facilitated the transformation of liberal arts colleges into comprehensive universities.

Finally, we are interested in the persistent effect from denomination competition on college quality. Estimates of DenomFraction on ranking is presented in Table 15. For top 100 research universities, DenomFraction in 1870 has a robust, positive effect on 2018 ranking. Interestingly, total endowment also has a lasting effect on rank, possibly through continuous investment in quality.

9 Conclusion

In this paper, we examine a previously unexplored origin of the private provision of human capital. We provide empirical evidence showing how historical episodes of religious competition have long-term benefits on education infrastructure and the production of human capital. Given the persistent US lead in higher education that would be un-eclipsed to this day and the consensus towards its significance for development, the origins of this "knowledge industry" is a understudied but salient episode of the development of the American economy.

We acknowledge potential concerns with endogeneity and pursue several strategies to alleviate them. Our structural estimates of school choice indicate that students in 19th century had strong preferences to attend colleges denominationally affiliated to their own church, this implies substantial gains to differentiation and illustrates an economic mechanism for our results.

A direction of future work is to model historical university competition in a framework which endogenizes decisions over entry, denominational affiliation, tuition setting, and quality determination. The model embeds a discrete choice setup within a sequential entry game where colleges decide whether to enter the market, then choose a denominational affiliation, taking into account consumer demand, the responses of other entering colleges, and the effect of affiliation choice on tuition and quality. This will allow us to conduct meaningful counterfactual simulations.

In summary, the unique absence of state sponsored religions in the United States led to a proliferation of religious denominations and factions in the 19th century. This historical fact had unintended and unplanned consequences for the growth of higher education.

Tables & Figures

Table 1 – DESCRIPTIVE STATISTICS

Denomination Share	1850		1860		1870		1890	
	mean	max	mean	max	mean	max	mean	max
Baptist	0.269	1.0	0.250	1.0	0.249	1.0	0.271	1.0
Christian	0.024	1.0	0.043	1.0	0.047	1.0	0.045	1.0
Dutch_Reformed	0.006	1.0	0.005	0.6	0.003	0.5	0.003	1.0
Episcopal	0.031	1.0	0.034	1.0	0.033	1.0	0.027	1.0
Quaker	0.010	0.7	0.007	0.6	0.004	1.0	0.005	0.3
German_Reformed	0.005	0.4	0.006	0.4	0.006	0.5	0.009	0.5
Jewish	0.000	0.1	0.000	0.1	0.000	0.2	0.001	0.1
Lutheran	0.022	1.0	0.025	0.7	0.030	1.0	0.049	1.0
Mennonite	0.001	0.1	0.000	0.0	0.000	0.0	0.001	0.4
Methodist	0.357	1.0	0.382	1.0	0.360	1.0	0.314	1.0
Moravian	0.005	0.4	0.001	1.0	0.000	0.0	0.000	0.2
Presbyterian	0.134	1.0	0.099	1.0	0.116	1.0	0.086	1.0
Catholic	0.065	1.0	0.080	1.0	0.098	1.0	0.083	1.0
Swedenborgian	0.000	0.0	0.000	0.1	0.000	0.0	0.000	0.1
Tunker	0.001	0.2	0.000	0.0	0.008	0.3	0.026	0.6
Union	0.015	1.0	0.025	0.8	0.000	0.0	0.000	0.0
Unitarian	0.002	0.2	0.002	0.2	0.001	0.2	0.002	0.2
Universalist	0.006	0.2	0.006	0.2	0.002	0.2	0.003	0.2
Congregational	0.026	0.7	0.030	1.0	0.031	1.0	0.040	1.0
Minor_Sects	0.020	1.0	0.000	0.0	0.000	0.0	0.007	0.4
Adventist	0.000	0.0	0.000	0.1	0.000	0.1	0.005	1.0
Mormon	0.000	0.0	0.005	1.0	0.010	1.0	0.011	1.0
Shaker	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.0
Spiritualist	0.000	0.0	0.000	0.0	0.000	0.0	0.000	0.1
Evangelical	0.000	0.0	0.000	0.0	0.003	0.4	0.009	0.3
Pentecostal	0.000	0.0	0.000	0.0	0.000	0.0	0.001	0.1

Notes: The table reports the average and maximum share of denomination accommodation at county level for each denomination.

Table 2 – Summary Statistics of Student & College Linked Data

Sample:	Religion reported			Missing religion		
	N	Mean	S.D.	N	Mean	S.D.
Panel A: Students						
Distance from home to college	2489	209.481	306.692	2632	191.109	332.641
Northeast	2703	0.79171	0.40615	2681	0.83849	0.36806
Midwest	2703	0.07362	0.26120	2681	0.06825	0.25223
South	2703	0.10691	0.30906	2681	0.08690	0.28175
Year Born	2419	38.2592	20.7082	2681	38.7116	18.4132
Urban	956	0.35041	0.47735	1221	0.45208	0.49790
Family Size	6	12.5	20.4230	3	10	0
Religious	2703	0.98446	0.12370	–	–	–
Presbyterian	2621	0.41739	0.49322	–	–	–
Congregational	2621	0.14803	0.35520	–	–	–
Episcopal	2621	0.12743	0.33352	–	–	–
Common parental occupations						
Minister	905	0.24640	0.43115	821	0.11571	0.32007
Farmer	905	0.13149	0.33812	821	0.07795	0.26826
Doctor	905	0.13038	0.33691	821	0.14129	0.34853
Common first occupations						
Minister	2680	0.51902	0.49973	2504	0.12985	0.28603
Lawyer	2680	0.08992	0.28612	2504	0.27236	0.44526
Teacher	2680	0.12350	0.32908	2504	0.15255	0.35963
Panel B: Colleges						
Tuition ($ per semester)	2234	58.8833	65.1926	2213	51.4629	20.5316
Same Religion	2285	0.55971	0.49650	–	–	–
# of instructors	2273	24.1011	25.2089	2208	28.4959	26.7092
Founding year	2786	1785.097	64.9371	2675	1777.313	68.86467
# of volumes in library	2203	53352.5	62578.5	2195	40890.1	57310.5
# of students	2267	277.567	222.370	2208	313.611	236.574

Notes: The table reports the mean, standard deviation, for two samples of students and colleges appearing in the student and college linked data.

Table 3 – Effect of Religious Share on College's Denominational Choice

College Affiliation	Baptist	Methodist	Presbyterian	Congregation	Episcopal
Baptist Share	0.5848***	0.1685	0.0345	0.0378	-0.0167
	(0.1807)	(0.2071)	(0.2149)	(0.1326)	(0.1455)
Methodist Share	-0.3168	0.4060*	-0.0202	0.1156	0.1554
	(0.1981)	(0.2271)	(0.2357)	(0.1454)	(0.1596)
Presbyterian Share	-0.0220	-0.0276	1.1545***	-0.0592	-0.2238
	(0.2455)	(0.2814)	(0.2920)	(0.1801)	(0.1977)
Congregation Share	-0.0844	-0.1520	-0.2519	1.5319***	0.0382
	(0.2301)	(0.2638)	(0.2737)	(0.1688)	(0.1853)
Episcopal Share	-0.1094	0.4928	-1.6199***	-0.3534	1.6388***
	(0.4532)	(0.5195)	(0.5391)	(0.3325)	(0.3650)
County level controls	Yes	Yes	Yes	Yes	Yes
Observations	230	230	230	230	230
R^2	.12	0.075	.16	.37	.13

Notes: The table reports linear estimates where the unit of observation is a college in 1860. Dependent variables are dummy variables that take a value of 1 when a college chooses denominational affiliation to be Baptist, Methodist, Presbyterian, Congregation, Episcopal and Christian respectively. Religious share of each of the six denominations are included, with share of own denomination as variable of interest. We control for a set of county level social economic conditions: total denominational accommodations, population, foreign population ratio, manufacturing output and a dummy for railroad access.

Table 4 – Effect of Religious Rivalry on College Founding: Cross Estimates

Dependent Variable:	Ln (Colleges)			
	(1)	(2)	(3)	(4)
Cross Section 1850				
DenomFraction	0.4034***	0.3554***	0.3505***	0.2015**
	(0.1343)	(0.1168)	(0.1181)	(0.0822)
N	1,615	1,615	1,615	1,615
R^2	.2	.22	.23	.27
Cross Section 1860				
DenomFraction	0.4536***	0.3778***	0.4141***	0.2014**
	(0.1481)	(0.1183)	(0.1207)	(0.0887)
N	2,073	2,073	2,073	2,073
R^2	.15	.19	.2	.24
Cross Section 1870				
DenomFraction	0.4172***	0.4051***	0.4135***	0.3379***
	(0.1282)	(0.1230)	(0.1136)	(0.0953)
N	2,289	2,289	2,289	2,289
R^2	.19	.22	.22	.24
Cross Section 1890				
DenomFraction	0.7318***	0.5687***	0.5295***	0.2196**
	(0.1725)	(0.1404)	(0.1330)	(0.0879)
N	2,779	2,799	2,799	2,799
R^2	.18	.21	.21	.25
State fixed effect	Yes	Yes	Yes	Yes
Denominational Controls	No	Yes	Yes	Yes
Geological Controls	No	No	Yes	Yes
Social Economic Controls Control	No	No	No	Yes

Notes: The table reports log-linear estimates where the unit of observation is a county. $DenomFraction$ is our constructed Herfindahl Index of denominational competition that equals $1-\sum_j Accommodation_j^2$, where *Accommodation* defines the sitting capacity of a denomination. Coefficient of $DenomFraction$ is reported, with clustered standard errors in brackets. County-level controls include religiosity measured by total sitting capacity in churches, farm productivity, distance to coast, distance to great lakes, distance to metropolitan cities, share of population corresponding to people in urban area, male, aged 5 to 18, white, literate, foreigners, access to railroad, miles of railroad, access to steam-boat navigated rivers, manufacturing output, manufacturing employment, manufacturing investment. The panel excludes counties that are only observed once or twice. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 5 – Effect of Religious Rivalry on College Founding: Panel Estimates

Dependent Variable:	Ln(Colleges)					
	(1)	(2)	(3)	(4)	(5)	(6)
DenomFraction	0.2907***	0.2984***	0.2965***	0.2247**	0.2131**	0.2235**
	(0.0767)	(0.0935)	(0.0934)	(0.0999)	(0.1024)	(0.1029)
Total Accommodations	–	0.0170***	0.0169***	0.0086	0.0073	0.009
		(0.0058)	(0.0058)	(0.0064)	(0.0069)	(0.0069)
County fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Denomination Controls	No	Yes	Yes	Yes	Yes	Yes
Farm Productivity	No	No	Yes	Yes	Yes	Yes
Demographics Control	No	No	No	Yes	Yes	Yes
Manufacturing & RR	No	No	No	No	Yes	Yes
Time trend	No	No	No	No	No	Yes
Observations	7,276	7,276	7,276	7,276	7,276	6,199
R^2	0.580	0.699	0.699	0.702	0.703	0.704
Oster's δ						1.802

Notes: The table reports log-linear estimates where the unit of observation is a county. *DenomFraction* is our constructed Herfindahl Index of denominational competition that equals $1 - \sum_j Accommodation_j^2$, where *Accommodation* defines the sitting capacity of a denomination. Coefficient of *DenomFraction* is reported, with clustered standard errors in brackets. County-level controls include religiosity measured by total sitting capacity in churches, share of population corresponding to people in urban area, male, aged 5 to 18, white, foreigners, farm productivity, access to railroad, miles of railroad, access to steam-boat navigated rivers, manufacturing output, manufacturing employment, manufacturing investment. The panel excludes counties that are only observed once or twice. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 6 – Testing Effect of Religious Rivalry on the Founding of Other Institutions

Dependent Variable:	Land Grants	Land Grants	Public School1850	Public College1890
DenomFraction	0.0264	0.0171	-1.3024***	0.0661
	(0.0169)	(0.0164)	(0.3274)	(0.0532)
State Colleges		0.3161***		
		(0.1110)		
County level controls	Yes	Yes	Yes	Yes
Denominational fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
Observations	2,600	2,600	1,600	2,600
Clusters	49	49	36	49
R^2	0.096	.14	.59	0.1

Notes: The table reports log-linear estimates where the unit of observation is a county. Column 1 & 2 dependent variables are the number of Land Grant Colleges founded by 1890. Column 3 dependent variable is the number of Public Schools reported in 1850 U.S census and column 4 dependent variable is the number of Public Colleges, the sum of Landgrant Colleges and State Chartered Colleges, existed in 1890. $DenomFraction$ is our constructed Herfindahl Index of denominational competition that equals $1 - \sum_j Accommodation_j^2$, where $Accommodation$ defines the sitting capacity of a denomination. Coefficient of $DenomFraction$ is reported, with clustered standard errors in brackets. $PrivateColleges$ is the number of non-Land Grant Colleges. County-level controls include total denomination accommodations, population, urban population, non-slave population, gender ratio, manufacturing investment, manufacturing employment and manufacturing output.

Table 7 – Reverse Causality test

Dependent Variable:	DenomFraction1890			
	All states	North	South	Midwest
Colleges1860	0.0016	-0.0002	0.0006	-0.0001
	(0.0011)	(0.0010)	(0.0023)	(0.0015)
DenomFraction1860	0.0707***	0.2971***	0.1512	0.0036
	(0.0238)	(0.0808)	(0.1080)	(0.0222)
Denominational fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
County level controls	Yes	Yes	Yes	Yes
Observations	1,873	546	341	662
Clusters	35	8	6	8
R^2	.69	.77	.62	.5

Notes: The table reports linear estimates where the unit of observation is a county. Dependent variable is DenomFraction of 1890. Independent variable of interest is $Colleges1860$, the log of number of colleges in 1860. $DenomFraction$ is our constructed Herfindahl Index of denominational competition that equals $1 - \sum_j Accommodation_j^2$, where $Accommodation$ defines the sitting capacity of a denomination. Coefficient of $Colleges1860$ is reported, with clustered standard errors in brackets. Column 1 include all states in 1860. Column 2 includes the Northern states: Maine,Pennsylvania, New Jersey, New York, New Hampshire, Connecticut, Massachusetts, and Rhodes Island.Column 3 includes the Southern States: Delaware,Maryland, South Carolina, Virginia, North Carolina, and Georgia. Column 4 includes Midwestern states: Iowa, Illinois, Michigan, Ohio,Indiana, Wisconsin, Minnesota, and Missouri. County-level controls include total denomination accommodations, population, urban population, non-slave population, gender ratio, manufacturing investment, manufacturing employment and manufacturing output.

Table 8 – DESCRIPTIVE STATISTICS

Variable	Revival towns		All towns	
	mean	sd	mean	sd
Total number of revivals	4.760	4.21		
Total number of revivals in 1831	1.696	2.00		
Population	2605.902	2190.36	2355.291	2077.14
Manufacturing share of population	0.037	0.02	0.036	0.02
Agriculture share of population	0.209	0.07	0.207	0.07
Commerce share of population	0.004	0.01	0.004	0.01
Total Area	85.720	88.57	79.711	93.94
Mean elevation	975.728	517.12	985.793	554.50

Notes: The table reports the mean and standard deviation of variables for NY towns with at least 1 revival meeting and for all NY towns. Population, manufacturing share of population, agriculture share of population and commerce share of population were collected from the U.S censuses of 1820. New York City and Albany was excluded from both samples.

Table 9 – Placebo Test of Instrumental Variable on Various Outcomes

	(1)	(2)	(3)	(4)	(5)
Dependent Variables:	College pre	Population	Manufacture	Commerce	Agriculture
Dist. to Rochester	-0.0605	-1692.87	236.50	163.97	-396.59
	(0.04)	(3065.54)	(436.10)	(153.26)	(314.45)
Population composition	Yes				
Township level Controls	Yes	Yes	Yes	Yes	Yes
Geological Controls	Yes	Yes	Yes	Yes	Yes
Observations	521	521	521	521	521
R^2	0.03	.66	.4	.17	.65

Notes: The table reports linear regressions. The dependent variables are log number of colleges built after 1830 in Column 1, Population in 1840 in Column 2, Manufacturing population in 1840 in Column 3, Commerce population in 1840 in Column 4 ,Agriculture population in 1840 in Column 5. An Observation is a New York town according to 1820 town boundary. Township controls include population in 1820, number of converts from revivals before 1830. Geological controls include distance to New York City, distance to Albany, distance to Canals and waterways, altitude and total area. Population composition controls for share of population corresponding to people in agriculture, commerce and manufacture. New York City and Rochester are excluded from both samples. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 10 – Revivalism and colleges: First Stage & Reduced Form

Dependent Variable:	First Stage				
	(1) Converts	(2) Revivals	(3) Converts	(4) Converts	(5) Colleges
Rainfall 1831	-20.364***	-0.326***	-21.95**	-22.377***	-0.006
	(5.862)	(0.074)	(8.718)	(6.689)	(0.004)
Dist. to Rochester	-0.2117*	-0.0030***	-0.2702*	-0.2617**	-0.145*
	(0.1138)	(0.0012)	(0.1517)	(0.1258)	(0.078)
Rainfall 1830			12.006		
			(11.852)		
Rainfall 1832			-9.562		
			(10.766)		
Dist. to Albany				-0.073	
				(0.058)	
Revivals before 1831	0.3157**	0.0026***	0.3184**	0.3163**	0.00002
	(0.1446)	(0.0008)	(0.1462)	(0.1446)	(0.00004)
Geological Controls	Yes	Yes	Yes	Yes	Yes
Township Level Controls	Yes	Yes	Yes	Yes	Yes
Observations	521	521	521	521	521
R^2	.27	.24	.27	.27	0.08

Notes: The table reports linear regressions. The dependent variable for the first stage is the number of revival meetings or number of converts in 1831. An Observation is a New York town according to 1820 town boundary. Township controls include population, share of population corresponding to people in agriculture, commerce and manufacture. Geological controls include distance to New York City, distance to Canals and waterways, altitude and total area. New York City and Rochester are excluded from both samples. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 11 – Revivalism and colleges: Second Stage

Dependent Variable: Ln(Colleges)	(1) IV-Robust	(2) IV-GMM	(3) OLS
Converts in 1831	0.0247*	0.0245*	0.0021
	(0.0137)	(0.0137)	(0.0027)
Geological Controls	Yes	Yes	Yes
Township Level Controls	Yes	Yes	Yes
Observations	521	521	522
R^2	.	.	0.080
LM-statistic	21.049	21.049	.
F-statistic	10.861	10.861	.
J-statistic	1.087	1.087	.

Notes: The table reports IV estimates. The dependent variable is the number of colleges in 1860 built after 1830 . An Observation is a New York town according to 1820 town boundary. Converts in 1831 equals the total number of converts from revival meetings from 1830-1832. Township controls include revivals prior to 1830, population in 1820, share of population corresponding to people in agriculture, commerce and manufacture in 1820. Geological controls include distance to New York City, distance to Canals and waterways, altitude and total area. New York City and Rochester are excluded from both samples. Kleibergen-Paap rk LM-statistic, Kleibergen-Paap F-statistic and Hansen J-statistic are reported. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 12 – College Choice, Conditional Logit Coefficients

	Parameter Estimates			
	(1)	(2)	(3)	(4)
Willingness-to-Pay ($)	77.943***	94.427***	334.60 ***	318.16***
	(28.03)	(27.37)	(27.98)	(0.0510)
Willingness-to-Travel (km)	239.31***	249.40***	231.39***	265.87***
	(64.12)	(-51.53)	(-42.71)	(-35.79)
College quality controls	No	Yes	No	Yes
County of college fixed effect	No	Yes	No	Yes
County of home fixed effect				
Observations	258768	250600	114552	110810

This table shows results from estimating the conditional logit model discussed in Section 7, with standard errors clustered at the county of home level. The sample consist of all student-college linked observations where the denomination of student is observed. Specification 1 includes only distance from home to college and an indicator variable for if the denomination of school matches that of the student. Specification 2 adds additional college characteristics such as tuition, the quality of school, coeducational status, etc. Specifications 3-5 differ in their choice of fixed effects. They include college county FE, county of college & home FE, college FE, respectively. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 13 – Short Run College Quality

Dependent Variable:	Students	Faulty	Grad school	Majors
DenomFraction	338.2515**	317.2102**	0.5983**	0.3941*
	(150.1081)	(124.0547)	(0.2764)	(0.2182)
State fixed effects	Yes	Yes	Yes	Yes
County level controls	Yes	Yes	Yes	Yes
Denominational Controls	Yes	Yes	Yes	Yes
Total students	No	Yes	Yes	Yes
Observations	228	232	421	421
R^2	.344757	.483913	.277071	.224979

Notes: The table reports OLS estimates. An observation is a private college. The dependent variable in Column 1 is the percent growth of total students from 1870 to 1890. The dependent variable in Column 2 is the percent growth of total faculty from 1870 to 1890. The dependent variable in Column 3 is per capita endowment in 1875, given that a college survived into 1895. The dependent variable in Column 4 is a dummy that equals 1 if a college had graduate students in 1910. The dependent varible in Column 5 is variable that equals 0 if a college in 1910 had no major, equals 1 if it had 1-5 majors and equals 2 if it had more than 5 majors. *DenomFraction* is our index of denomination competition in 1870.County-level controls include religiosity measured by total sitting capacity in churches, share of population corresponding to people in urban area, male, aged 5 to 18, white, foreigners, farm productivity, access to railroad, miles of railroad, access to steamboat navigated rivers, manufacturing output, manufacturing employment, manufacturing investment. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 14 – Long Run College Quality: Becoming Research University

Dependent Variable:		Research University			
Grad Program	0.1694***	0.1182**	0.1063**	0.1021**	0.1042**
	(0.0496)	(0.0483)	(0.0448)	(0.0442)	(0.0461)
Total Students		0.0001***	-0.0000	0.0000	0.0000
		(0.0000)	(0.0001)	(0.0001)	(0.0001)
Total Faculty			0.0019***	0.0025***	0.0025***
			(0.0007)	(0.0008)	(0.0008)
Total Endowment				-0.0002***	-0.0002***
				(0.0001)	(0.0001)
Range of Major					-0.0293
					(0.0884)
County level controls	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes
Denominational fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	239	239	239	221	221
R^2	.474931	.528888	.55681	.622937	.623367

Notes: The table reports OLS estimates. An observation is a private college in 1910 that survived to today. The dependent variable is a dummy variable that equals 1 if a private college is a research university (ranked $\leq$ 100) today. Total Endowment is total endowment in thousands of dollars in 1910. Range of majors is variable that equals 0 if a college in 1910 had no major, equals 1 if it had 1-5 majors and equals 2 if it had more than 5 majors. County-level controls include religiosity measured by total sitting capacity in churches, share of population corresponding to people in urban area, male, aged 5 to 18, white, foreigners, farm productivity, access to railroad, miles of railroad, access to steam-boat navigated rivers, manufacturing output, manufacturing employment, manufacturing investment. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 15 – Long Run College Quality: Ranking as a Research University

Dependent Variable:	US News Ranking			
DenomFraction	61.4185*	63.3741**	62.8741*	67.3437**
	(30.8939)	(31.5570)	(31.6958)	(30.7391)
Total Endowment	0.0255***	0.0290**	0.0262**	0.0214*
	(0.0063)	(0.0116)	(0.0125)	(0.0122)
Total Students		-0.0020	-0.0045	-0.0033
		(0.0057)	(0.0071)	(0.0068)
Total Faculty			0.0416	0.0745
			(0.0677)	(0.0669)
Range of Majors				-25.6432**
				(10.3943)
Population	Yes	Yes	Yes	Yes
Observations	82	82	82	82
R^2	.251402	.252616	.256311	.312132

Notes: The table reports OLS estimates. An observation is a college, private or public, in 1910 that survived to today and ranked by US News in 2018. The dependent variable is ranking, with 1 being the highest rank. *DenomFraction* is our index of denomination competition in 1870. Total students and total faculty are the total number of undergrad students and faculty in 1910. Range of majors is variable that equals 0 if a college in 1910 had no major, equals 1 if it had 1-5 majors and equals 2 if it had more than 5 majors. Total Endowment is total endowment in thousands of dollars in 1910. Population in 1990 is controlled for. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Figure 2 – DENOMINATION COMPETITION IN 1860

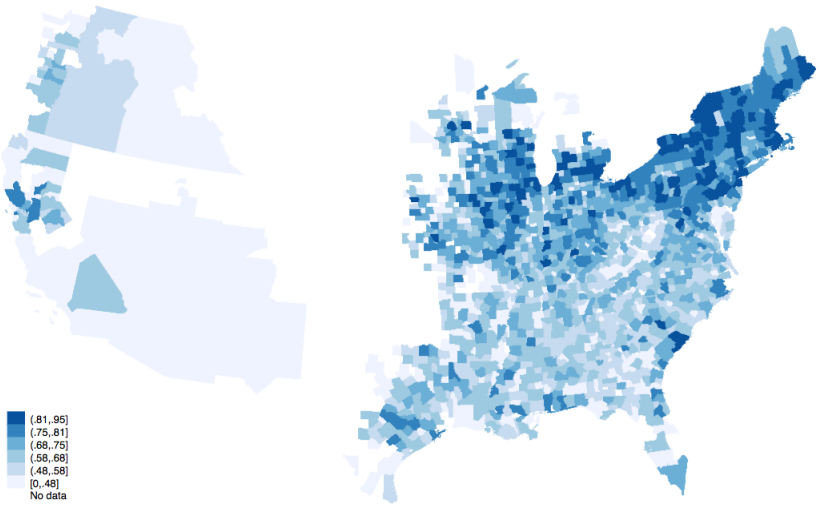


Figure 1 – DENOMINATION COMPETITION IN 1850

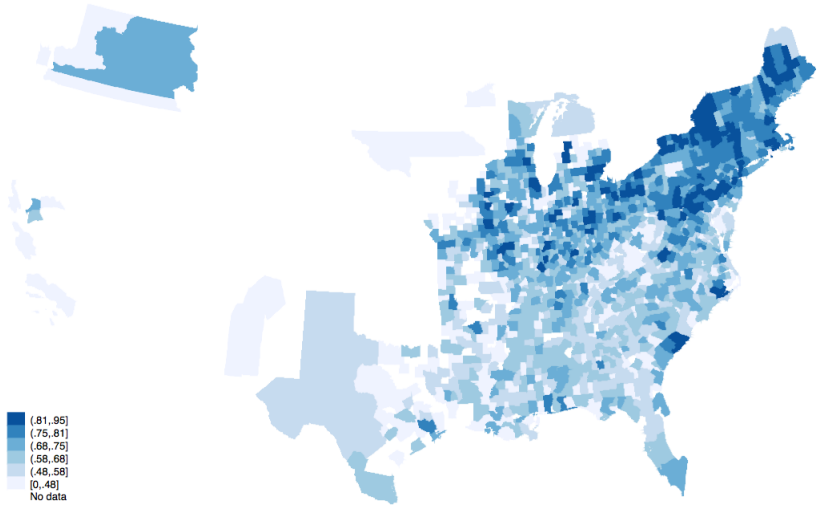


Figure 3 – DENOMINATION COMPETITION IN 1870

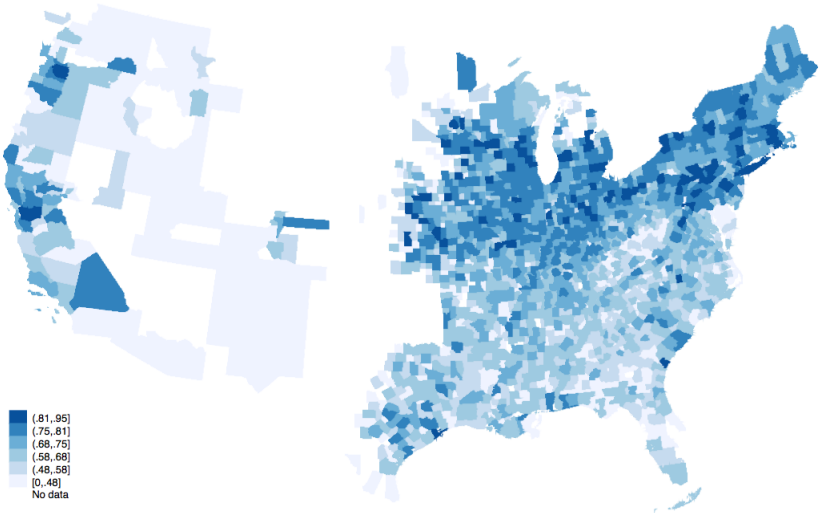


Figure 4 – DENOMINATION COMPETITION IN 1890

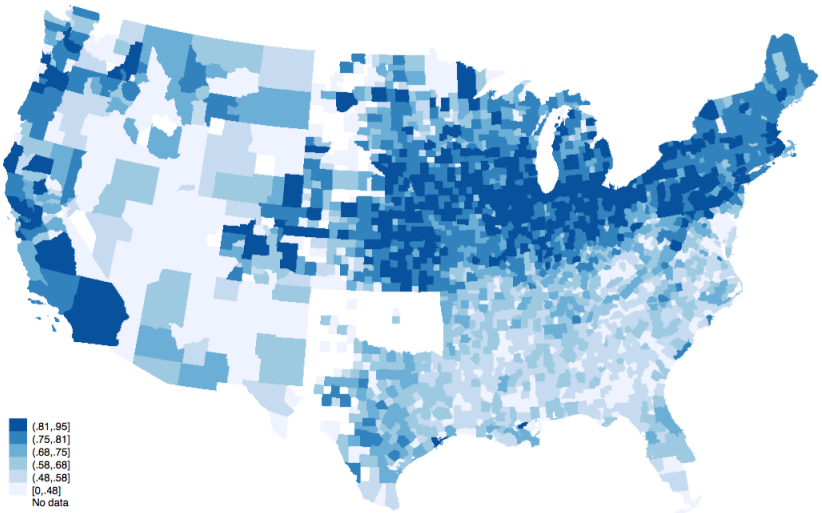


Figure 5 – Colin Burke

Shelby College. Surviving until at least the Civil War, this institution began some college work in 1841 after reorganization under Episcopal auspices. It may have lost the strict denominational influence it had in a reorganization in the mid–1850s. It was considered a failure by Tewksbury.

Figure 6 – College Expansion

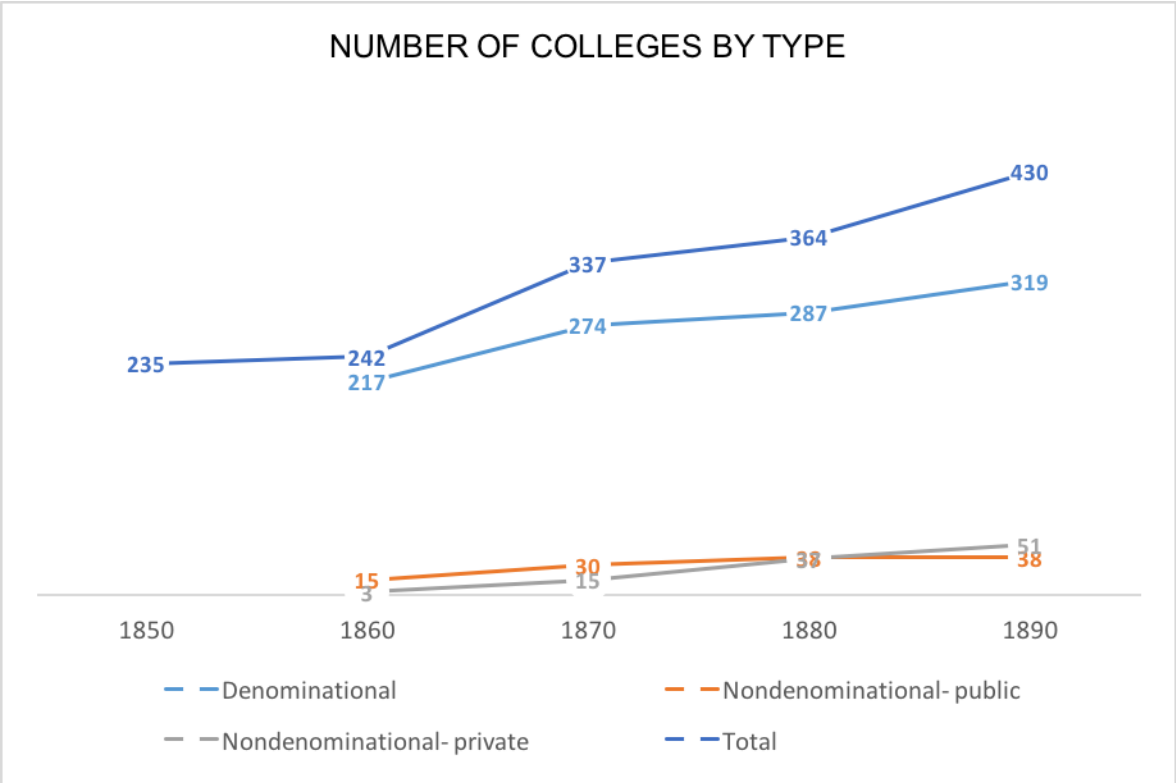
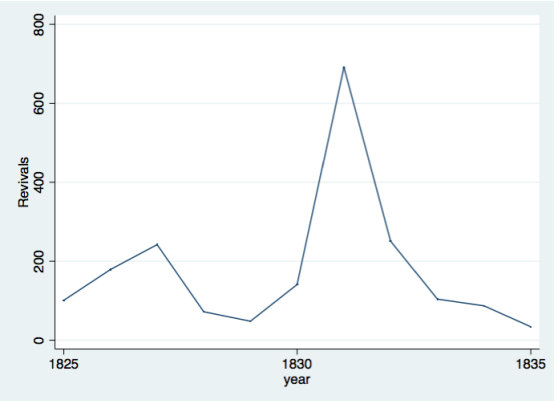
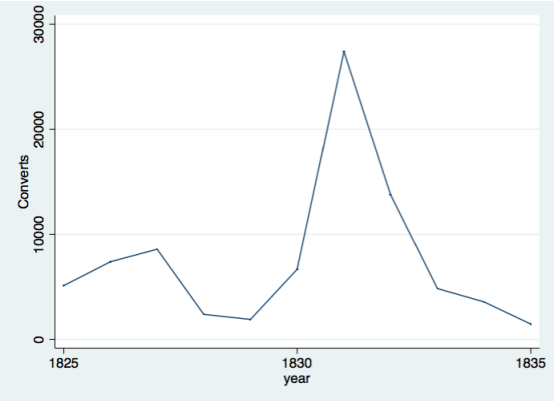


Figure 7 – Revivalism Activity Across Time

(a) No. of Revivals by Year



(a) No. of Converts by Year



(c) No. of Counties involved in Revivals by year

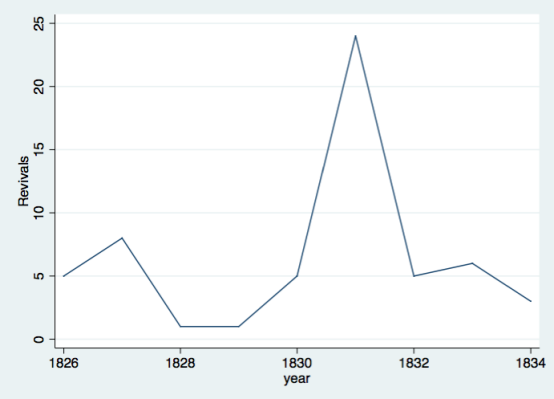
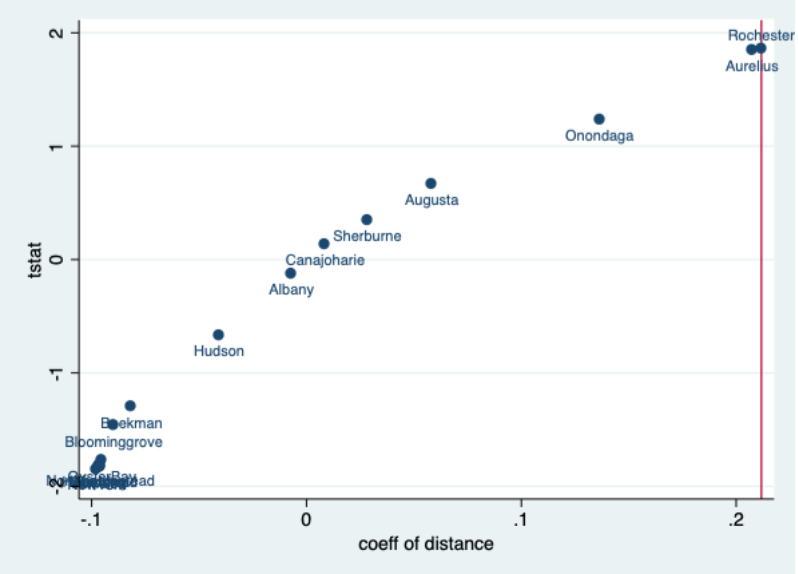


Figure (a), Figure (b) and Figure (c) present the trend of revival activity measured by total number of revival meetings, total number of converts and total number of affected counties respectively.

Figure 8 – Placebo First Stage

(a) Distance Coefficient on Converts



(b) Distance Coefficient on Revivals

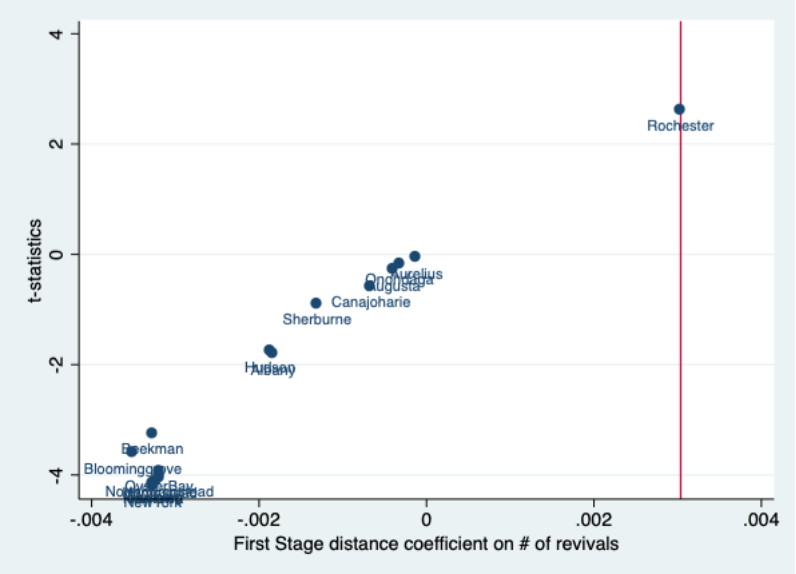
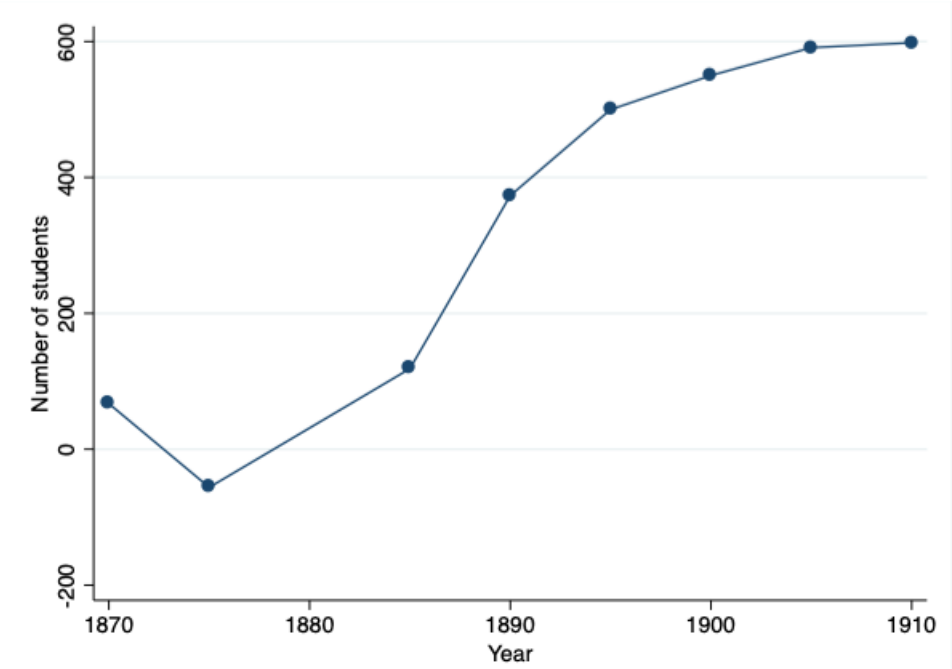


Figure a and Figure b draw comparisons between the first stage coefficients from Columns 1&2 of Table 10, respectively, and the coefficients from the same specification, but estimated with the distance from 16 New York towns with population greater than 10,000 by 1820, as opposed to Rochester. Red vertical lines indicate the distance to Rochester coefficients from Columns 1&2 of Table 10. Dots represent the negative of first stage coefficients for 16 New York towns and Rochester

Figure 9 – Short-Run Effects on Institutional Quality

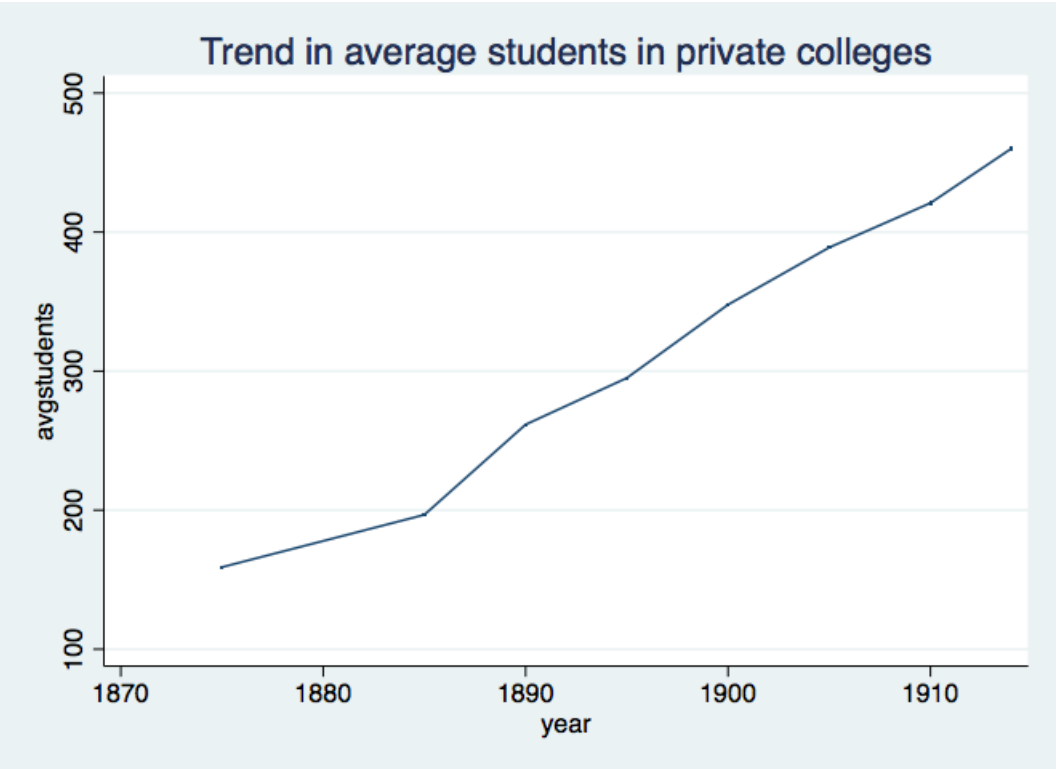
(a) Effect of Religious Diversity on total students over time



(b) Effect of DenomFraction on total faculty over time

Notes:

Figure 10 – Growth of total students in private colleges



References

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, **91** 1369–1401.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, **8**.

Andrews, M. (2017). *Fuel of Interest and Fire of Genius: Essays on the Economic History of Innovation*. Ph.D. thesis, University of Iowa.

Atack, J. (2013). On the use of geographic information systems in economic history: The american transportation revolution revisited. *The Journal of Economic History*, **73** 313–338.

Bai, Y. and Kung, J. K.-s. (2011). Climate shocks and sino-nomadic conflict. *The Review of Economics and Statistics*, **93** 970–981.

Becker, M. K. M., Gary S. (2003). *Social Economics: Market Behavior in a Social Environment*. Belknap Press.

Becker, S. O. and Woessmann, L. (2009). Was weber wrong? a human capital theory of protestant economic history. *The Quarterly Journal of Economics*, **124** 531–596.

Bisin, A. and Verdier, T. (2000). "beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, **115** 955–988.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, **99** 977–1009.

Bubolz, R. G. . J. A. (2000). College as it was in the mid-nineteenth century. In *The American College in the Nineteenth Century* (R. Geiger, ed.). Vanderbilt University Press, Nashville, 80–90.

Burke, C. (1982). *American collegiate populations-a test of the traditional view*. New York University Press.

Cagé, J. (2015). The economics of the african media. In *The Oxford Handbook of Africa and Economics: Volume 2: Policies and Practices* (C. Monga and J. Y. Lin, eds.). Oxford University Press, Oxford, 605–25.

Cantoni, D. (2013). The economic effects of the protestant reformation: Testing the weber hypothesis in the german lands.

Cantoni, D. and Yuchtman, N. (2012). Medieval universities, legal institutions, and the commercial revolution. Tech. rep., National Bureau of Economic Research.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, **69** 1127–1160.

Chaney, E. (2013). Revolt on the nile: Economic shocks, religion, and political power. *Econometrica*, **81** 2033–2053.

Chaudhary, L. and Rubin, J. (2011). "reading, writing, and religion: Institutions and human capital formation". *Journal of Comparative Economics*, **39** 17–33.

Church, R. L. and Sedlak, M. W. (1997). *The Antebellum College and Academy*. Simon & Schuster Custom Publ., Needham Heights, MA.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics,*, **112** 1203–1250.

Findlay, J. (2000). Agency, denominations, and the western colleges, 1830-1860. In *The American College in the Nineteenth Century* (R. Geiger, ed.). Vanderbilt University Press, Nashville, 115–126.

Finke, R. and Stark, R. (1992). *Churching of America 1776-1990*. Rutgers University Press.

French, J. (1860). *Gazetteer of the State of New York*. Syracuse, N.Y., R.P. Smith.

Geiger, R. (2000). The era of multipurpose colleges in american higher education. In *The American College in the Nineteenth Century* (R. Geiger, ed.). Vanderbilt University Press, Nashville, 127–152.

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

Geruso, M. and Spears, D. (2017). Neighborhood sanitation and infant mortality. *NEBR Working Paper*.

Goetzmann, W. H. (2009). *Beyond the Revolution: A History of American Thought from Paine to Pragmatism*. Basic Books.

Goldin, C. and Katz, L. F. (2009). *Why the United States Led in Education: Lessons from Secondary School Expansion, 1910 to 1940*. Cambridge University Press. NBER WP No. 6144.

Goldin, C. and Katz, L. F. (2010). *The race between education and technology*. Harvard University Press.

Gordon, T. F. (1836). *Gazetteer of the State of New York*. Philadelphia, PA: Printed for author.

Haines, M. R. (2010). Historical, demographic, economic, and social data: The united states, 1790-2002. URL https://doi.org/10.3886/ICPSR02896.v3.

Hammond, J. L. (2007). Revivals in new york and ohio, 1825-1835. **101**.

Iannaccone, . B. E., L. (2008). Economics of religion. In *The New Palgrave Dictionary of Economics* (S. Durlauf and L. Blume, eds.). Palgrave Macmillan, New York.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, **100** 271–291.

Iyer, S., Velu, C. and Weeks, M. (2014). Divine competition: Religious organisations and service provision in india. URL https://doi.org/10.17863/CAM.4953.

Jha, S. (2013). Trade, institutions and ethnic tolerance: Evidence from south asia. *American Political Science Review*, **107** 806–832.

Johnson, C. D. (2008). "sectarian nation": Religious diversity in antebelllum america. *OHA Magazine of History*, **22** 14–18.

Kuran, T. (2014). *The Long Divergence: How Islamic Law Held Back the Middle East*. Princeton University Press.

Kuran, T. (2016). Islam and economic performance: Historical and contemporary links.

Mazzeo, M. J. (2002). Competitive outcomes in product-differentiated oligopoly. *Review of Economics and Statistics*, **84** 716–728.

McBride, M. (2008). Religious pluralism and religious participation: A game theoretic analysis. *American Journal of Sociology*, **114** 77–106.

Mccleary, R. and Pesina, J. d. J. (2012). Protestantism and human capital in guatemala.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, **3** 303–328.

Mokyr, J. (2009). *Enlightened Economy*. Yale University Press.

Montgomery, J. D. (2003). A formalization and test of the religious economies model. *American Economic Review*, **68** 782–809.

Potts, D. B. (1971). American colleges in the nineteenth century: From localism to denominationalism. *History of Education quarterly*, **11** 363–380.

Prummer, A. and Siedlarek, J.-P. (2014). Institutions and the preservation of cultural traits.

Putnam, C. D. E., Robert D. (2000). *American Grace: How Religion Divides and Unites Us*. New York University Press.

Rogers, R. L. (2010). The urban threshold and the second great awakening: Revivalism in new york state, 1825-1835. **49** 694–709.

Seim, K. (2001). *Spatial differentiation and market structure: The video retail industry*. Ph.D. thesis, Yale University.

Spafford, H. G. (1824). *Gazetteer of the State of New York*. Albany, NY: B.D.Packard.

Squicciarini, M. P. and Voigtländer, N. (2014). Human capital and industrialization: Evidence from the age of enlightenment. Tech. rep., National Bureau of Economic Research.

Stone, L. (1968). Literacy and education in england 1640-1900. *The past and present society* 69–139.

Tewksbury, D. G. (1932). *The founding of American Colleges and Universities before the Civil War*. Ph.D. thesis, Teachers College, Columbia University.

Woodberry, R. (2012). The missionary roots of liberal democracy. *American Political Science Review*, **106** 224–274.