NORTHWESTERN UNIVERSITY

Essays in Economics History of American Higher Education

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR Of PHILOSPHY

Field of Economics

By

Yiling Zhao

EVANSTON, ILLINOIS

December 2020

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

Essays in Economics History of American Higher Education

Yiling Zhao

This thesis contains three chapters studying the evolution of the American higher education landscape, the different forces that shaped their organization, and how they, in turn, influenced human capital accumulation. The chapters are organized into three time periods: 1850-1900, 1910-1940, and 1980-2010. In the first chapter, joint with Heyu Xiong, we investigate the founding of American colleges in the latter half of the 19th century. During this period, America experienced a significant expansion in its college infrastructure. By 1890, more institutions of higher learning existed in the United States than in all of Europe. In this paper, we study the role of denominational competition in the market provision of higher education. Specifically, we argue that denominational affiliation facilitated greater enthusiasm to build colleges through gains from differentiation. We document that nearly all colleges established in this period had denominational roots or origins. Empirical analysis using a panel of US counties reveals a robust positive relationship between religious fragmentation and the number of colleges established locally. The results using an IV strategy show that areas that became more religiously diverse experienced greater subsequent growth in the number of colleges. We conclude by estimating a model of school choice and show that students exhibited a strong preference to attend same-denominational colleges in terms of willingness-to-pay and willingness-to-travel. Religious diversity softened the extent of tuition competition between institutions and precipitated an "excess" entry of schools.

In the second chapter, joint with Mike Andrews, we study a particular aspect of the land-grant colleges: the institutionalization of Home Economics. The land-grant colleges offered practical majors that contrasted with the historical practice of a liberal arts curriculum provided by private colleges. Home Economics was such a practical field for women. Even though Home Economics built on women's traditional vocation as homemakers, we argue that it served as a gateway to

bring women into science. Specifically, we propose that women came to dominate life sciences and chemistry because they were exposed to these subjects in large numbers through college home economics curricula in the late nineteenth and early twentieth centuries. As more women learned these subjects and entered these fields, they increasingly became known as womens subjects, reducing social stigma. We document three sets of facts. First, the science fields in which women constitute a majority today are the same fields that entered home economics curricula during the late nineteenth and early twentieth centuries, a time of rapid increase in college enrollment. Second, we use data from the Report of the Commissioner of Education to show that colleges with larger home economics programs had more women enrolled in science. Since home economics programs were often designed to train farmers wives, we verify these results using an instrumental variable based on a colleges land grant funding towards agriculture. Finally, we use data from historical college yearbooks to show that, at an individual level, women who major in home economics are more likely to double major in the sciences than are women who major in other subjects, such as music and education.

The third chapter connects integrally to the second chapter, as I explore how gender norms can influence contemporary college major choices. I explore the hypothesis in the setting of the computer science major from 1980 to 2010. In this period, women's representation increased in STEM fields overall, yet the percentage of women in computer science declined. I take advantage of computer science's shifting departmental affiliation from mathematics to the traditionally masculine domain of engineering. Using a novel panel dataset on the university hierarchy from 1980 to 2010, I found that the percentage of women earning computer science bachelor's degrees decreased when the computer science department moved from colleges of liberal arts and sciences to engineering schools. I document that computer science's increasing affiliation with engineering school had more to do with following the trend of higher-ranked programs than reflecting a change in the curriculum.

ACKNOWLEDGEMENTS

I am indebted to my committee chair Joel Mokyr, for introducing me to economic history, commenting on every draft of my paper, and teaching me to be a much better writer. I thank my advisor Carola Frydman for her frank feedback and hard questions, which inspired many parts of this dissertation. I thank my advisor Joe Ferrie for his American Economic History class and his deep reservoir of historical and technical knowledge.

I am fortunate to have worked with some of my brilliant peers, Stephanie Johnson, Heyu Xiong, and Mike Andrews. I have learned a great deal through collaborating with them.

I am also grateful to my graduate school friends, particularly Nico Inostroza, Luca Bittarello, and Melanie Xue, for always being there for me. I thank my husband Jack, for his unconditional support in the past five years. I would not have survived graduate school without them.

Last but not the least, I thank my undergraduate advisor, George Neumann, for encouraging me to go to graduate school. He discovered my interest in economic history and advised me on my very first research project. Unfortunately, he passed away in 2015 and didn't have a chance to read any of the chapters. I hope I have made him proud.

TABLE OF CONTENTS

Acknowledgments 5

List of Tables 9

List of Figures 12

Chapter 1: Sectarian Competition and the Market Provision of Human Capital 15

- 1.1 Introduction 17
- 1.2 Related Literature & Contributions 22
- 1.3 Historical Background 24
- 1.4 Data 27
  - 1.4.1 Denominations & Supply of Colleges 32
- 1.5 County-Level Empirical Analysis 33
  - 1.5.1 Cross Section 33
  - 1.5.2 Panel Specification 35
  - 1.5.3 Placebo Checks 37
  - 1.5.4 Additional Results & Robustness 38
- 1.6 The Second Great Awakening 39
  - 1.6.1 Data & Identification Strategy 41
  - 1.6.2 Empirical Results 43

1.7 Student Preferences & School Choice 46
1.8 Conclusion 50

Chapter 2: Home Economics and Women's gateway to science 71
2.1 Introduction 71
2.2 Historical Background 74
2.3 Data 76
2.4 Cross Section Analysis 79
  2.4.1 Land Grant Sample 79
  2.4.2 Robustness Checks 80
  2.4.3 Agriculture Instrument 81
2.5 Panel Analysis with Student yearbooks 84
2.6 Mechanism 86
2.7 Conclusion 87

Chapter 3: Engineering the gender gap: fall of women's share in computer science . . . 100
3.1 Introduction 100
3.2 Historical background 108
  3.2.1 From big irons to minicomputers 108
  3.2.2 From programmer to software engineer 111
  3.2.3 From math to engineering 113
  3.2.4 Gender and Race 115
3.3 Literature & Contribution 118
3.4 Data 121

3.4.1 Department affiliation data 121

3.4.2 Graduation data 124

3.4.3 Some Illustrative Graphs 125

3.4.4 Comparison between the sample and the population 130

3.4.5 Determinants of Engineering Affiliation 133

3.5 Main analysis and results 138

3.5.1 Panel Specification 139

3.5.2 Discussion & Interpretation 141

3.6 Conclusion 146

References 166

Appendix A: Chapter1 168

Appendix B: Chapter2 169

Appendix C: Chapter3 171

LIST OF TABLES

1.1 Descriptive Statistics: Denomination Market Share 52
1.2 Effect of Religious Share on College's Denominational Choice 53
1.3 Effect of Religious Rivalry on College Founding: Cross Estimates 54
1.4 Effect of Religious Rivalry on College Founding: Panel Estimates 55
1.5 Testing Effect of Religious Rivalry on the Founding of Other Institutions 55
1.6 Robustness: Excluding Regions 56
1.7 Effect of Religious Rivalry on College Founding: State-Level Estimates 56
1.8 Effect of Religious Rivalry on College Founding: Alternative Treatment Def. . . . 56
1.9 Descriptive Statistics: New York State Towns 57
1.10 Placebo Test of Instrumental Variable on Various Outcomes 57
1.11 Revivalism and colleges: First Stage & Reduced Form 58
1.12 Revivalism and colleges: Second Stage 58
1.13 Student Micro-Data Summary Statistics 59
1.14 College Choice, Conditional Logit Coefficients 60
1.15 Short Run College Quality 61
1.16 Long Run College Quality 62

2.1 Spill over to science : land grant universities 89

2.2 The effect of funding sources on program sizes 90
2.3 Testing spill-over effect of other majors on science 90
2.4 Spill over to science : First Stage & Reduced Form 91
2.5 Placebo Test of Instrumental Variable on Various Outcomes 91
2.6 Spill over to science : Second Stage 92
2.7 Yearbook Data Summary Statistics 92
2.8 Panel Estimates from Yearbooks Sample 93
2.9 Cornell University: Men in Home Ec and Science 93

3.1 Determinants of the switch 138
3.2 Engineering Affiliation Effect 141
3.3 Determinant of Engineering Affiliation 147
3.4 Summary Statistics of students and college characteristics in 2010 148
3.6 Frequency of science terms mentioned in bibliography title 149
3.5 Home Economics bibliography: volumes and mean published year 149
3.7 Effect of offering Home Economics on women's participation in sciences 150
3.8 Engineering Affiliation Effect on Master's 151
3.10 CS-Engineering Gap effect 152
3.9 Effect of opening engineering program on women's share 152

B.1 Spill over to science: full sample 169
B.2 Cp, R-squared and Actions along the sequence of Lasso Algorithm 170

C.1 Major choice in Computer Science 190

C.2 Effect of engineering syllabus on expected enjoyment from coursework 190

LIST OF FIGURES

1.1 College Expansion 63
1.2 Denominational Competition: 1850-1890 64
1.3 College Expansion by Institutional Type 65
1.4 Secularization of higher education over time 66
1.5 Revivalism Activity Across Time 67
1.6 Placebo First Stage 68
1.7 Growth of total students in private colleges 69
1.8 Denomination Market share & affiliated college 70

2.1 Distribution of majors by institution type 94
2.2 Number of Degrees in Home Ec conferred 95
2.3 Home Ec Trends from the yearbooks sample 96
2.4 Share of women in science 97
2.5 Gender ratio in science 97
2.6 Gender ratio in science by treatment group 98
2.7 Value of farm equipment and agriculture program size, 1910 98
2.8 Home Ec and agriculture program sizes, 1910 99
2.9 Cornell University Hotel Management first year courses 99

3.1 Growth in computer related occupations: 1970-2010 109

3.2 Women's share by year and occupation 110

3.3 Programmer, Software developer & Software Engineering 113

3.4 White women in math, physical science and engineering 116

3.5 White women's representation in math, EE & CS 118

3.6 Computer Science Departmental Affiliation 122

3.7 Indiana University trend 127

3.8 Northwestern University trend 128

3.9 Clemson University trend 129

3.10 Fraction of women in CS by affiliation/subfield 131

3.11 Fraction of women in CS programs that "never","always", or at "some" point– . . . 132

3.12 Top 50 engineering school as trendsetters 136

3.13 Engineering affiliation and graduate engineering program ranking 137

3.14 CS Engineering gender gap & drop in CS's women share 144

A.1 Denomination Market share & affiliated college 168

C.1 Computer peripheral operators working on the mainframe IBM360 171

C.2 Distribution of home economics bibliography by subfield and year 172

C.3 Fraction of women in computer industry 174

C.4 Asian women's representation in math, EE & CS 175

C.5 Foreign students' representation in CS at Master's level 176

C.6 Computer Science Departmental Affiliation 177

C.7 Trend in the fraction of women earning bachelor's degrees 178

C.8 Rise of popular majors for women since 1970 179

C.9 Fraction of women earning bachelor’s degrees in math, CS and engineering 179

C.10 Fraction of women earning master’s degrees in math, CS and engineering 180

C.11 Share of men and women earning bachelor’s degrees in CS 180

C.12 Number of men and women earning bachelor’s degrees in CS 181

C.13 Trends in total CS programs and CS graduates 181

C.14 Total colleges and CS programs in sample 182

C.15 Correlation between CS and Engineering gender ratio 182

C.16 Sample Course Catalog 183

C.17 Degree requirement 184

C.18 Fraction of women in various science subjects 184

C.19 Fraction of women in various engineering subjects 185

C.20 Double Major Distribution by gender 186

C.21 Syllabus A 188

C.22 Syllabus B 189

CHAPTER 1

SECTARIAN COMPETITION AND THE MARKET PROVISION OF HUMAN CAPITAL

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

1.1 Introduction

In the United States, the closing decades of the 19th century witnessed a significant expansion in the scope and landscape of higher education. From 1820 to 1859, 225 private universities were built, and over the next 40 years, an additional 348 were founded (Goldin and Katz, 1999). By 1881, Ohio alone had 43 institutions for a population of 3 million people. In contrast, England had four universities serving a population nearly eight times that size (Goetzmann, 2009).[1] The high density of private universities and their remarkable quality would become an enduring characteristic of the American higher education system. However, where did this growth in educational infrastructure originate? What sustained such a seemingly disproportionate level of investment?

This article studies the economic and competitive forces that 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 the 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 tuition competition amongst colleges and precipitated an "excess" number 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 as crucial to its success and diffusion. However, with notable exceptions, this existing literature has focused on elementary and secondary education. Largely omitted is a discussion of higher education prior to 1890, when, paradoxically, denominational and private interests played a substantially more nuanced role. By exploring the origins and foundations of the "knowledge industry", this

---

[1] Similarly, France had 16 colleges.

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 small, as in all nations. However, 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 capital or literacy (Squicciarini and Voigtländer, 2014), and universities themselves played a remarkable role in facilitating the Commercial and Industrial Revolutions (Cantoni and Yuchtman, 2012; Mokyr, 2009). Furthermore, Card (2001) documents that even proximity to colleges impacts educational attainment decisions.[2] The striking implication of Card (2001) is that the equilibrium distribution of universities has 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's degree-granting institutions in the 19th century US. For each institution, we document the exact geographic location, land-grant status, cost of tuition and board, resources provided, level of endowment, and, importantly, denominational affiliation if stated. Using geographic information, we link these institutions back to the underlying markets in which they were initially built.

By 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 that includes detailed accounts of religious composition and membership. Specifically, we collect data on the total church accommodations belonging to each denomination within the county and their respective market shares. From this information, we compute a measure of a county's religious market concentration or fractionalization.

---

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

The first part of our empirical analysis reveals a positive and significant relationship between local religious diversity and the establishment of colleges, both in the cross section and over time. Our panel estimates relate the changes in the number of colleges within a county to the changes in the degree of religious fragmentation. A back-of-the-envelope calculation suggests that there would have been approximately 21 percent fewer colleges in 1890 if the US had been 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 controlling for an extensive set of confounding factors.

In particular, we control for county fixed effects, denomination fixed effects, population, urbanization, industrialization, education, the gender ratio, land productivity, agricultural output, geographic conditions, and railway or canal access. Denomination fixed effects show that our results are not driven by the presence of any single denomination, while county fixed effects account for unobservable heterogeneity, which is time invariant.

A key concern with this analysis, however, is that religious diversity could be related to time-varying unobserved factors, such as the demand or preference for education, which also affect college development. To address this, we conduct falsification checks. Placebo analysis shows that religious diversity had no positive impact on the number of high schools or land-grant colleges. The relationship exists *only* between religious competition and the supply of private denominational colleges. To exclude the possibility of reverse causality bias, we also show that the spatial distribution of colleges in 1860 does not predict religious diversity in 1890. Overall, both findings help us rule out competing explanations.

To identify the causal effect of religious competition, we adopt an instrumental variable strategy that exploits the onset and diffusion of the Second Great Awakening. The Second Great Awakening was a historical episode in the 19th century during which the United States underwent a resurgence in religious activity and experienced unprecedented expansion in church membership. The Second Great Awakening was composed of numerous religious revivals that gave rise to new denominations and altered the religious composition of affected localities. The intensity and frequency of revivals in a given location depended on the influence of prominent and persuasive pastors nearby, whose impact could spill over to surrounding communities.

This aspect of revivalism motivates our instrumental variable strategy. We focus on revivalism in New York State, where Pastor Finney, an influential religious leader based in Rochester, was active. Using township-level data, we instrument for revival intensity using the distance to Rochester. We confirm that towns closer to Rochester experienced more revival activities throughout the period. We show that revivalism translated into greater college growth after the Second Great Awakening through the mediating channel of increased religious fragmentation.

An extensive literature, beginning with the pioneering work of Max Weber, has 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. Comparatively, 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.[3]

Drawing from the industrial organization literature on differentiated products and entry, we attempt to disentangle the underlying mechanisms specific to our context. Starting a college in the 19th century US was a precarious proposition that carried a significant risk of failure. In this competitive environment, we posit that denominational differentiation provided a mechanism to insulate entrants from onerous competition on tuition 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 college denomination to be a dimension for horizontal differentiation such 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 with a denominational affiliation that matches their own; two, colleges recruit students locally.

---

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

Without fully formalizing the model, we assess the strength of these parameters that govern the underlying incentives for and gains from 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 about the demand side of colleges in the 19th century. We construct a unique student-level dataset assembled from college directories, alumni records and biographies that contains information on student characteristics and the college attended. We estimate a conditional logit discrete choice model and diagnose how various factors, such as religious affiliation, distance, and tuition, affect observed choice.

We find that students strongly preferred to attend a college with the denomination that matched their own. Our preferred estimates indicate that students were willing to travel up to 239-249 kilometers farther and pay an additional \$77-\$94 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. 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 remainder of the paper is organized as follows. Section 1.2 conducts a literature review and discusses our relative contributions. Section 1.3 provides a brief history of denominational involvement in higher education. Section 1.4 explains our contribution to the data and the sources of our datasets. Section 1.5 describes our main empirical specifications and reports the results. Section 1.6 presents the analyses based on the Second Great Awakening, which utilizes quasi-experimental variation in New York State. Section 1.7 estimates an empirical model of school choice that allows us to recover the preferences for attending colleges of the students' own denominations. Finally, Section 1.8 concludes the paper.

1.2 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), Becker et al. (2016), Mccleary and Pesina (2012), Cantoni (2013) and Cag (2015) suggests 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 a specific religion or denomination matter. This is echoed in evidence from developing settings. Examples include Geruso and Spears (2017), which explores the Hindu–Muslim cleavage's relevance for sanitation habits, Chaudhary and Rubin (2011) discuss the relationship among reading, writing and religion in colonial India. Other notable recent papers include Kuran (2016), Kuran (2014) and Chaney (2013).[4]

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

---

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

argued that competition benefited the consumers of religion and constrained the extent of rent extraction by religious authorities (Iyer, 2016).[5]

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).

However, the existing literature has been primarily concerned with the effect of religious competition on religious practices, whether religious participation or the shaping of religious customs; less well understood are 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 conflict.

In this domain, the closest paper to our own is Iyer et al. (2014), which provides evidence of religious competition and cooperation in contemporary India by examining religious and nonreligious service provision. To the best of 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 the economic implications of diversity. Studies have typically emphasized the economic costs of diversity. Easterly and Levine (1997) show that ethnic diversity adversely affects public policies associated with economic growth, such as the black market, low provision of infrastructure, and low levels of education. Alesina et al. (2003) find 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 primarily focused on ethnic and linguistic fractionalization; our paper explores the direct effect of religious diversity. Although the existing literature regards 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.

---

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

Finally, 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 explicitly quantify the preference for religious homophily in the context of 19th century higher education.

1.3 Historical Background

The connection between American higher education and religion dates back to the colonial period. Nearly all of the colonial colleges had denominational affiliations: Harvard and Yale were congregational, the College of William & Mary and Columbia were Anglican, and Brown was Episcopal, just to name a few.[6]

University services were intertwined with religious functions. In addition to providing a formalized curriculum of classical education, colleges trained and supplied future ministers.[7] Since denominations differed in their interpretation of the Bible, to prepare students for religious vocations, colleges catered to specific denominations, and denominational differences in university education truly mattered. The history 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 founding of Yale, as some evidence suggested, was to supply orthodox ministers to Connecticut churches (Tewksbury, 1932). Ministerial work remained a promising career path throughout the Antebellum period. More than half of Harvard graduates became ministers for the sixty years following its founding. Nearly three-fourths of the graduates of Yale became ministers for the first twelve years after its founding

---

[6] Two colleges, Princeton and the University of Pennsylvania, were officially nonsectarian but nevertheless were primarily influenced by Presbyterianism and the Church of England, respectively.

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

(Tewksbury, 1932). A steady 17 percent of Princeton graduates became ministers from 1824 to 1854 (Bubolz, 2000).

Denominational involvement in higher education continued and intensified in the 19th century. Prior to the 19th century, the landscape of higher education in the United States was fairly static. Growth was minimal, with fewer than 50 institutions founded between 1638 and 1819. However, the pace of expansion accelerated considerably after 1820, which ushered in a period of unprecedented expansion. Using the data from Goldin and Katz (1999), Figure 1.1 shows the remarkable rise in the number of college foundings after 1820. The growth in colleges vastly exceeded the growth in population, with a total of five colleges per million people in 1790, rising to 16 per million in 1880. By 1880, the US had five times as many colleges as the entire continent of Europe. This was the most overbuilt system of higher education in existence.

Nearly all colleges built during this time had affiliations with a specific denomination or church. Religious denominations played a vital role in procuring economic support for affiliated colleges. In some cases, a denomination motivated local donors. Potts (1971) stated that the official or verbal sanction from a religious organization legitimized the endeavor and encouraged local financial support. Since colleges founded in the early 19th century located in rural towns far away from major population centers and recruited students from areas within a 50-mile radius (Church and Sedlak, 1997), denominational influence on local economic support mattered.

In other cases, denominations played a more active role in university finances. They formed agencies and organizations that specialized in college fundraising. For instance, by order of the Indiana Conference (a Methodist society), all Methodist ministers in Indiana were agents for Indiana Asbury University, with the responsibility 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 Congregationalist societies had organizations analogous to those of the Methodists. Some societies even planned based on their agents' visiting schedules to avoid exhausting donors' interest in giv-

---

[8] "The best method of collecting funds, that is, to not 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)

ing. Agents' salaries, which varied with their performance, were deducted proportionally from the sums collected. According to John Peck, an agent of Shurtleff College (Baptist), approximately 40 to 60 percent of funds gathered reached college coffers (Findlay, 2000). The denominational network became a financial resource and allowed institutions to secure resources beyond their immediate surroundings.

The ability to fundraise beyond a local network provided denominational colleges with a source of competitive advantage. This became especially true as access to formal credit markets was difficult given the nonprofit disposition of the projects. As a result, denominational colleges enjoyed a greater likelihood of survival. Ninety percent of Protestant colleges founded before 1860 survived through the Civil War, while only thirty-seven percent of nondenominational colleges and seventy-four percent of state colleges did. (Burke, 1982). In the absence of a nondenominational 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, rather than faculties (Burke, 1982)..

A contributing factor was the intense periods of Protestant religious revivals that took place during the early 19th century. This is commonly referred to as the Second Great Awakening. The movement introduced innovations such as camp meetings (an outdoor religious gathering usually held in a rural area), which helped to enroll new practitioners. Active denominations such as Baptist, Methodist and Presbyterian grew into the dominant positions and saw a marked increase in their membership. A number of new denominations such as Seventh-day Adventist, Disciples of Christ, Church of Christ, etc. were born in the midst of the Second Great Awakening.

These newer denominations entered into the competition for higher education. In the 1810s, the three oldest denominations, Congregationalists, Episcopalians and Presbyterians, were associated with 70% of all colleges; by the time of the Civil War, this share had 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 out of the need to educate priests and more from the concern about losing the allegiance of young men educated by rival denominations. Newer denominations, such as Methodist and Baptists, who were anti-formalists in nature and opposed existing seminaries (Johnson, 2008), were affiliated with one-third of all colleges by 1860.

Competition among different denominations stimulated denominational college building and the frenzy to establish schools. This competitive relationship is probably best illustrated in the Old Northwest. In the state of Ohio alone, 43 institutions were founded in this period, including Kenyon (1824), Western Reserve (1826), Oberlin (1833), Denison (1831), and Marietta (1835) (Geiger, 2000; Goetzmann, 2009).

Denominational influence in higher education diminished following the end of the 19th century, after which industrial philanthropy took on a more prominent role in college development. Prior to the 1900s, nondenominational fundraising had been sporadic and unpredictable (Geiger, 1997). Since the early 20th century, educational fundraising has become more professionalized through organized advancement and development programs (Chan, 2016). Large-scale philanthropy and private donations from newly wealthy industrialists and their families crowded out the relative importance of denominational patronage. Substantial magnitudes of alumni donations also became, for the first time, a recurrent and dependable source of income (Geiger, 1997). Consequently, the comparative advantage of denominational affiliation declined, and from the 1900s onward, many formerly religious institutions began undergoing a process of secularization.

1.4 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 over 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 US Census began in 1790, and the compilation of religious data did not begin until 1850. A balanced sample of 1,632 counties included in the panel analysis.[9]

From 1850 to 1890, census enumerators gathered facts concerning the number of churches, their locations, and their seating accommodations. The information is broken down by denominational affiliation. In 1850, the census identified 18 principal denominations. To the best of our ability, we standardize denomination categories over time to account for differences in the granularity of definitions. Nevertheless, by 1890, there are 24 denominations, which reflects the religious growth and innovations that characterized the period. Table 1.1 reports the average and maximum share of denominational accommodation at the county level for each denomination.

The Baptists and Methodists dominated throughout the decades we focus on, representing over half of the religious share. They were followed by three distant competitors: Presbyterian, Episcopalian, and Congregationalist. Although these three enjoyed the prime status of state-supported religions in the colonial era, their influence waned as their authority passed. Lutherans and Catholics thrived as considerable flows of immigrants entered 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 across 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 Herfindahl-Hirschman index of the share 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 is not very concentrated. Specifically, the denominational competition in county c is:

$$DenomFraction_c = 1 - \sum_i s_{ic}^2 \tag{1.1}$$

---

[9] We harmonize the data according to the 1850 county boundaries. The steps to adjust census data to a base year were proposed by Hornbeck (2010)

where s_{ic} is the share of denomination i's accommodations as fraction of total church accommodations in county c. Figures 1.2 show the spatial distribution of denominational competition from 1850 to 1890: it was high in the Northeast and Midwest but also in Gold Rush California and in frontier states such as Colorado. Beyond regional disparities, there were also sizable differences across counties in each state, allowing identification from within-state variation.[10]

College Information and Location: Compiling information from a number of sources, we build an original dataset of 19th century US colleges. In 1850, the US 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; therefore, 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 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 courses and compiled those that did into a list.

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

The *Report of Commissioners of Education* provided rather detailed information on colleges since the 1870s on an annual basis. In these reports, the commissioners conducted censuses of in-

---

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

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

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

stitutions of higher education on a voluntary basis through surveys. Extensive information was collected, including name, location, denomination affiliation, founding date, endowment, land value, library volume, enrollment, expenses, tuition and board. All the information was self-reported by each college. To address and mitigate problems associated with nonresponse, we interpolate missing data by filling in information provided in adjacent years. Ultimately, we are able to assemble a rather complete set of colleges for 1870 and 1890.

Figure 1.3 plots the total number of colleges in our compiled dataset from 1850 to 1890. We are able to identify denominational colleges from 1860 to 1890. We divide the category of nondenominational colleges 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 secularization of existing denominational colleges. The new colleges built in this time period were overwhelmingly denominational.

Student Micro Data & College Choices: To the best of our knowledge, little is known about the demand side of colleges in the 19th century. Because of the lack of linked administrative data dating from this period, there is a scarcity of information on who attended colleges, their 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 the Antebellum college students from Colin Burke. He surveyed a large array of related materials held in the Library of Congress, the libraries 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 students and their educational paths.

The original dataset contains approximately 12,000 students, with their name, hometown zip code, college destination, and known denominational affiliation. Occupation, father's occupation and extra curriculum were recorded whenever available. After omitting observations with missing data, we retained approximately 3,000 students with a full set of covariates. Summary statistics of all students and colleges appearing in the sample are provided in Table 1.13.

The first observation of interest is that students on average traveled a relatively short distance, 200 kilometers, from their hometown to attend college. This reinforces the notion that colleges served local and isolated markets. Therefore, it is sensible 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 that the tuition charged by colleges was relatively low. Higher education in the 19th century, even among elite institutions, is best characterized as a buyer's market. The pool of applicants was small, and admission was not competitive. Colleges such as Columbia and Harvard regularly struggled to fill their fall classes and advertised steadily to do so through August or September, just days from the semester's start.[13] The strong competitive forces exerted downward pressure on tuition.

Finally, we observe that while the majority of students attended same-denomination colleges, cross-denomination enrollment was nontrivial. This suggests that, consistent with our reading of the historical narratives and school charters, denominational fit or compliance was rarely, if ever, a de facto requirement for admission or attendance.

Historical Controls: We gather county economic and demographic characteristics covering the period 1850-1890 from the US Census. Haines (2010) provides decadal, county-level data on manufacturing, and 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.

---

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

1.4.1 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 1.2 shows that the denominational composition within a county plays a significant role in driving the observed supply. We provide results for the six denominations that had the largest total accommodations. A unit of observation is a college in 1860. The table presents an OLS regression of a dummy for whether a college is affiliated with a certain denomination (Baptist in column 1, Methodist in column 2, Presbyterian in column 3, etc.) on the market shares of those denominations while controlling for geographic and socioeconomic controls. Given the construction of the dependent variable, the coefficient can be interpreted as the marginal effect of a denomination's market share on the likelihood of a college's affiliation 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 accommodations in a county, the more likely a Baptist college is. 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 that a 10 percentage point increase in the Episcopalian share increases the likelihood of an Episcopalian affiliation by 16 percentage points.

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 his or her own denomination and that fundraising depends heavily on the local network, a college's likelihood of adopting a denomination should increase with its share of the local population. This provides preliminary evidence on the salience of affiliation choices.

We also show that denominational influence in higher education declined following the end of the 19th century. Figure 1.4 illustrates that the number of colleges founded as nondenominational institutions and the rate of secularization amongst previously denominational institutions rose sharply from 1890 onwards.

1.5 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 the corresponding instrumental-variable solution is presented in the next section.

1.5.1 Cross Section

The cross-sectional estimation equation is:

$$log(\# \; of \; colleges_c) = \beta DenomFraction_c + \sum^{j} \gamma_j Accommodation_{j,c} + \lambda_s + \sigma X_c + \epsilon_c, \quad (1.2)$$

where the outcome is the 1 + the total number of colleges in county c in logarithms, $Accommodation_{j,c}$ defines the seating capacity of 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, denomination fixed effects, geographic controls, and socioeconomic controls. Definitions of these socioeconomic controls and their sources are given in the appendix.

We control for state fixed effects and cluster standard errors at the state level to control for unobservables that affect settlement patterns for both denominations and colleges that are geographically fixed. This is intended to capture sharp regional differences in habitability, the availability of public funds, social structure (the practice of slavery) and some degree of cultural attitude. Column 1 in Table 1.3 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 markets, and these shares may be strongly correlated with religious competition. Since denominations vary in their attitudes towards higher education, we address this by controlling for denomination shares and including dummies for each of the largest five denominations that take 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 identified only from variation in the composition of religious bodies as opposed to the presence of any single denomination. Column 2 in Table 1.3 shows that the results are robust to controlling in various ways for the dominance of denominations.

Geographic controls include county area, agricultural productivity, the (Euclidean) distance to the coast, distances to great lakes, and distances to major metropolitan cities. Including geographic controls is important because they affect the likelihood of settlement for both denominations and candidate colleges. Distance to the coast, waterways and metropolitan cities proxy for market access, which could increase religious diversity because market access is known to increase trade, population and income. Column 3 of Table 1.3 shows estimates after controlling for geographic conditions.

We consider three subcategories of socioeconomic controls: demographic characteristics, industrialization, and transportation infrastructure. The first category consists of the shares of population corresponding to people in urban areas, out-of-state migrants, foreigners, males, whites, and those 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 steamboat-navigated rivers. Column 4 in Table 1.3 shows the results of this specification. A large body of literature has confirmed that railroads and canal networks have significant implications for banking, urbanization and industrialization. Many anecdotal studies have also suggested railroad connections as a key factor considered when college founders selected candidate locations. The robustness of the results to controlling for transportation is especially reassuring.

1.5.2 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 to which religious fragmentation varies at different points in time to investigate the effect of changes in religious diversity on changes in the number of colleges. In contrast to the previous analysis, our estimates are identified not by regional variation in the cross section but variation at a given location over time.

The estimating equation is:

$$log(\#\, of\ colleges_{c,t}) = \beta DenomFraction_{c,t} + \sum^{j} \gamma_j Accommodation_{j,c,t} + \delta_c + \sigma X_{c,t} + \epsilon_{c,t} \qquad (1.3)$$

Formally, we replace the state fixed effects with county fixed effects λ_c because the US incorporated 14 new states from 1850 to 1890, and county boundaries changed over time. We perform the analysis in a balanced panel and harmonize the data according to the 1850 county boundaries.[14] Our panel dataset contains 6,487 observations and 1632 unique counties.

---

[14] The steps to adjust census data to a base year were proposed by Hornbeck (2010). We adopt the harmonizing code developed by Perlman that follows the Hornbeck (2010) method.

Column 1 in Table 1.4 shows the results of the baseline panel regression. Standard errors are clustered at the county level to account for spatial autocorrelation over time. The inclusion of the county fixed effect allows us to flexibly address any unobserved, time-invariant county characteristic that may differentially affect growth or attract settlement.

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

Following the same line of reasoning, access to steamboat-navigated rivers is excluded. We expect few changes in canal building from 1850 to 1890, as most canals were completed by the early nineteenth century. In addition, all socioeconomic controls and railroad infrastructure controls are maintained, and estimates from including socioeconomic controls are shown in Columns 4 & 5. Finally, by including decadal fixed effects, the treatment effect of religious fragmentation is only identified from variation within a county and decade.

By comparing the changes in point estimates and R-squared across specifications in Table 1.4, we can assess the sensitivity of our results to omitted variable bias. Following the steps outlined in Oster (2019), we test for the significance of unobservable components. The test statistic suggests that selection on unobservables would need to be at least twice as important as selection on observables to yield a coefficient of zero.

To put the effect size into context, we consider the counterfactual where the US had only a single denomination. The average county-level denominational fragmentation equals 0.6675 in 1890, using panel estimates in column 6 Table 1.4, the US would have had 0.6675 X 0.32 = 21% fewer colleges in the complete absence of denominational diversity.

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

1.5.3 Placebo Checks

Our panel estimates identify the effect of denominational diversity on college building under the assumption that unobservable factors that might simultaneously affect colleges and religious diversity are *time invariant*. In this section, we discuss possible sources of bias and potential threats to identification.

Several factors could contaminate the OLS estimates. In particular, it is natural to suspect that religious diversity may be correlated with time-varying unobservables that affect either the demand for or supply of higher education. For example, religious diversity may reflect cultural tolerance and a propensity to adopt new ideas, which could result in greater demand for higher education. On the other hand, religious diversity could be correlated with economic potential, which is related to the cost of college construction.

In the preceding section, we attempt to address these competing explanations by directly controlling for them. The controls ruled out differences in population, demographics, manufacturing, transportation networks, land productivity, migration, religious norms, etc. as possible confounders. Nevertheless, we acknowledge that these precautions are not exhaustive, and the threat of omitted variable bias remains.

To address the remaining concerns, we conduct a series of placebo checks designed to validate our proposed mechanism. Specifically, we test whether denomination fragmentation correlates with the establishment of educational institutions such as state or land-grant universities and high schools, where the denominational forces likely would not apply. However, if demand for higher education were driving the underlying results, then OLS regression on those institutions would still yield positive results.

To further alleviate concerns about selection, we also examine whether denominational diversity predicts college growth after 1900, when denominational influence in college building declined and the importance of denominational affiliation subsided. To implement these tests, we collect location data on existing state and land-grant universities that were designated to be built in 1890; high schools established by 1850; and colleges built after 1900. We run log-linear regressions on these public institutions using 1890 religious and socioeconomic data.

Table 1.5 shows the results of this placebo test. Column 1 reports the null effect of denomination fragmentation on land-grant colleges in 1890. Column 4 describes 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. If our results on private colleges are driven by the overall demand for education, then we might expect a positive relationship between denominational diversity and public high school provision. Financial support for secondary education differs from that of universities. Since funds for public high school were collected from real estate taxes regardless of whether a family had children to send to school, a key determinant of public school demand is a homogeneous population (Goldin and Katz, 2009). Such homogeneity spans the dimensions of income, parental education level and, what matters to us, religion. Therefore, we expect a negative relationship between the religious diversity index and the density of public high schools. Column 3 in Table 1.5 is consistent with this idea.[15]

1.5.4 Additional Results & Robustness

Market & sample definitions: Our analysis defines a county as a college market. We do this because the county is the smallest unit at which we can disaggregate religious composition data over our study period. However, a nonnegligible segment of the student population attended college out of county. This induces an imprecision in the sense that the religious diversity we calculate may not correspond to the relevant market considered. To account for this, we relax our market

---

[15] While the estimates from the public high school regression validate the identifying assumption, it introduces another endogeneity 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. This could attenuate the results.

definition and conduct our analysis at the state level. States serve as a reasonable upper bound on the market definition of a college.[16] The state-level analysis is shown in Table 1.7. The larger estimates suggest significant attenuation bias at the county level.

To address concerns that any particular region drives the findings, we also estimate our main models while dropping individual census regions: the Northeast, Midwest, South, and West. Table 1.6 presents the results for this set of robustness exercises. The results are qualitatively similar across all subsamples. We take the robustness of the positive relationship across the different models for both outcomes and samples to be compelling evidence that our estimated impacts are not driven by any particular group of states but rather reflect a general pattern.

Alternative treatment construction: Our main explanatory variable is the inverse Herfindahl-Hirschman index of market concentration based on county-wide denominational seat shares. This index combines variation both in terms of which denominations were active in the county and the number of church seats provided by each denomination. The latter measures local religious diversity to the extent that church accommodation approximates church membership. However, when churches are not seated to capacity, measurement error will undoubtedly ensue.

As a robustness check, we employ another measure of religious diversity that relies on variation in the extensive margin alone. Namely, we replicate our main analysis by relating college growth to the number of denominations operating in the county. The results are presented in Table 1.8. Consistent with our main findings, we show that the number of denominations present in a county-year is positively associated with the number of colleges constructed. The coefficients imply that each additional denomination in a county leads to an 2.69 percentage point increase in the number of colleges.

1.6 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 un-

---

[16] To date, the majority of students attend colleges within their state of residence.

fortunate shortcoming of our earlier 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 processes. Although we control for correlates to the best of our ability, the issue of joint determination in our research question is of critical concern. The remaining unobserved characteristics are serious threats because they might bias, or in the worst case invalidate, the hypothesized relationship between religious diversity and higher education.[17] 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 within.

As discussed in the historical background section, the early nineteenth century witnessed the spread of religious revivalism commonly referred to as the Second Great Awakening. The Second Great Awakening was a Protestant religious movement that began in the late eighteenth century and continued through the first third of the nineteenth century. 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 state-sponsored religion. Hence, religious proselytizing began in the colonial period and accelerated in the early national period, rising to a fever pitch during the Second Great Awakening. During this period, membership in evangelical denominations grew twice as rapidly as population growth (Finke and Stark, 1992), and 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.[18] The revival events reshuffled the local religious composition and gave rise to entirely new denominations.

---

[17] 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 explicitly control for migration, but counties with pronounced diversity might also be characterized by a more liberal socioeconomic environment open to new cultures and ideas.

[18] Formalists consisted of elite Presbyterian, Reformed Dutch, and Congregationalists; representatives of Anti-Formalists were Methodists and Baptists.

These camp meetings led extraordinary numbers of people to convert through an enthusiastic style of preaching and audience participation. Characterized as emotional, enthusiastic, and appealing to the supernatural, the content of revival events had little connection to education.

Given the vital role of opinion leaders in driving the success of revival events, the location and frequency of revivals depended on the itinerary of influential pastors. Building from this observation, we instrument for revival events with proximity to influential pastors and study how revival events lead to subsequent college establishments through the mediating channel of religious diversity. We utilize several datasets from New York State from 1820 to 1860 to perform the causal analysis.

1.6.1 Data & Identification Strategy

Our data on revivalism come from Hammond (2007), who collected information from religious newspapers on religious revivals in New York between 1825 and 1835. For each instance, he recorded the location, year, and number of people affected.[19]

Hence, our empirical analysis is confined to a decade-long period in one state. However, the choice of period and location is not arbitrary. New York, 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 in many comparable regions.

Furthermore, the period covers the influential Rochester Revival (9/3/1830-3/3/1831), which is considered a significant point in the 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 activity

---

[19] 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), Christian Advocate (New York, 1826-1835), Evangelical Magazine and Gospel Advocate (Utica, 1830-1831), and Visitant (Utica, 1825-1827). Unfortunately, denomination was not specified in each revival entry.

reached its arguable zenith. Figure 1.5 shows the yearly trend in revivalism activity by the total number of events and total number of converts. 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 acquire digitized demographic data from Rogers (2010). The original data were taken from the US census of 1820, and the census towns were merged with revival locations referenced by 19th-century gazetteers (French, 1860; Gordon, 1836; Spafford, 1824). A total of 521 townships are available for study, with variables on population size, share of population corresponding to manufacturing, agriculture and commerce, and total area. We geocode the towns to add transportation and geographic controls. Distances to canals and waterways in 1831 are computed using shape files from Atack (2013). Geographic controls include distance to major cities and mean elevation.

We construct religious fragmentation at the township level using the seating accommodations of each denomination from the New York State Religious Censuses (Benson and Silbey, 2006). We construct the inverse Herfindahl index of religious fragmentation in 1845, the first year New York State conducted its religious census.

Table 1.9 presents descriptive statistics of NY towns. Columns 1&2 report the mean and standard deviation of variables for towns with at least one revival meeting (revival towns); Columns 3&4 report those for all other NY towns.[20] Revival towns are slightly more populous and industrialized than the average New York town.

Identification Strategy:

While the expressed purpose of revival events had little connection to higher education per se, 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. In our setting, revivals cause two outcomes of interest: denominational fragmentation as the intermediate outcome and establishment of colleges

---

[20] New York City and Albany were excluded.

as the final outcome of interest. We apply mediation analysis with a single instrument, according to Dippel et al. (2019).

We focus on incidents of revivals in 1831 specifically, as that year constituted a major shock in the flow of revival activities, as shown in Figure 1.5.

We leverage the fact that Charles Finney's Rochester Revival inspired follow-up 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 outward. Therefore, we construct the Euclidean (straight line) distance from each town in our sample to Rochester to capture this element of spatial diffusion.

$$DenomFrac_{j,1845} = \gamma_1 \cdot Revivalism_{j,1831} + \sigma \cdot Distance\ to\ Rochester_j + \pi X_{j,1820} + \upsilon_j \quad (1.4)$$

$$log(\#\ of\ colleges_{j,1860}) = \beta \cdot Denom\hat{F}rac_{j,1845} + \gamma_2 \cdot Revivalism_{j,1831} + \epsilon_j, \quad (1.5)$$

where $Denom\hat{F}rac_{j,1845}$ is the estimated value of $DenomFrac_{j,1845}$ in the first stage. The endogenous treatment, $Revivalism_{j,1831}$, is the incidences of revivalism activities in town j in 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 instrument in the first stage is $Distance\ to\ Rochester_j$, the Euclidean distance between Rochester and town j. Finally, X_j is a vector of town-level sociodemographic and geographic characteristics as well as the history of revival activities in the town in 1820.

1.6.2 Empirical Results

First Stage & Placebo Tests:

Before turning to the 2SLS results, we examine the first-stage relationship in our township sample. Columns 2&3 of Table 1.11 report the coefficient on the instruments in predicting the total number of revivals from 1825-1835 and denomination fragmentation in 1845. The relevance assumption is satisfied, as towns farther away from Rochester experienced fewer revival meetings and became less religiously fragmented. 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 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. It is possible that the relevance of distance to Rochester is driven by the broader effect of market access rather than the actions of Pastor Finney. To cope with this possibility, we replicate our baseline first-stage regression for all New York towns with populations greater than 10,000 by 1820.

Figure 1.6 shows that the coefficient for Rochester lies at the top end of the distribution in both graphs. Only in the case of Rochester does distance exhibit a significant effect on 1831 revival activity. Rochester's relationship to revival intensity dominates every other large New York town in both magnitude and statistical significance. This evidence supports the singular importance of distance from Rochester, which is consistent with the narrative of Pastor Finney.

Finally, revival activities could have had a long-term impact on economic growth that correlates with demand for colleges. We show in Columns 2-5 of Table 1.10 that no significant relationship exists between distance to Rochester and growth in population or population in manufacturing, commerce or agriculture. Column 1 also documents that towns close to Rochester were not more likely to have colleges before 1830. This addresses any concern that distance to Rochester was correlated with preexisting demand for education. To summarize, the distance to Rochester provides a plausible means to obtain variation in 1831 revivals that is exogenous to other determinants of higher education.

IV Results:

Now, we quantify the marginal effect of denominational fragmentation on college establishments in the second stage. Table 1.12 presents the IV estimates in 2SLS (columns 1-3). The results are robust to controlling for geographic and preexisting demographic conditions. They are also robust to controlling for revivals, suggesting that revivals led to college establishment through the mediate outcome, denomination fragmentation. Our instruments proved to be relevant and strong: the F statistic is 19 before controlling for revivals and 9.5 after controlling for revivals. The second-stage results are comparable to the panel estimates.

The IV estimates are larger in size and more significant than the OLS estimates, which are reported in Column 3 of Table 1.12. The discrepancy between the IV and OLS estimates indicates 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 revivals with respect to educational investment. This accords with the narrative evidence that indicates that revivals were spiritually motivated and largely took place in rural communities that possibly placed less emphasis on higher education. 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 revival enthusiasm, and yet one is more suitable for building a college campus, while the others have natural settings more suitable for revivals. The difference between the OLS and IV estimates could also be due to attenuation bias induced by measurement error in the 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 compared to the 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 makes it readily apparent what the treatment is. This allows us to more easily assess the plausibility of the identification strategy. Whereas it was difficult to consider the exclusion 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. Ultimately, this allows us to make progress toward actual causal estimates of the effect of religious diversity on higher education.

1.7 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 that underlie these findings. We regard 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 exactly cater to the preferences of consumers. To the degree that the underlying taste in the population is heterogeneous, this will be reflected in the increased provision of varieties within a decentralized market. Colleges can avoid competing on price by maximally differentiating themselves along the religious spectrum. Importantly, the gains to differentiation from these standard Hotelling channels allow 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 as follows: 1) colleges serve local markets, and 2) consumers exhibit a preference for schools with denominational affiliations that match their own and regard colleges with the same denomination as more substitutable than colleges with different denominations. 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 colleges affiliated with the same denomination as their own?

For this purpose, we utilize a sample of linked 19th century student-college data where college choice and student demographics are jointly observed. The sources of these data are described in detail in the Data section. It contains rich demographic information about the students, and we merge 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 the distance students traveled to attend their respective college of choice.

We measure students' preference to attend schools with the same denomination by estimating a discrete choice demand model that uses 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 the 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:

$$U_{ij} > U_{ik}, \forall\, k \in \{1, ..., J\} \; and \; k \neq j \tag{1.6}$$

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 the unobserved idiosyncratic preference of student i for school j:

$$U_{ij} = X_{ij}\beta + \epsilon_{ij} \tag{1.7}$$

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 allows us to recover the utility parameters.

Several variables comprise the baseline X_{ij} vector, starting with an indicator variable for whether 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), the size of the school in terms of total enrollment, and the quality of the school as measured by faculty size, the volume of books in the library, and the founding date of the college. We also estimate alternative specifications where we include 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, which is a proxy for the unobserved labor market conditions or career opportunities at the location of college, which could have driven admission.

We conduct our analysis on two samples of students: 1) the subsample of students whose denominational backgrounds were recorded and 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 shown in Table 1.13, this group of students differs on several key observable characteristics from the overall population of students. Thus, the estimates based on this group are likely not representative of the true preference parameters and are more akin to an upper bound. With the full sample of students, we conservatively assume that every student whose denomination information was not recorded is nonreligious and did not attend a college of their own denomination. This biases the results against our expectation and provides a conservative lower bound on the religious preference of the overall student population.

The baseline estimates are reported in Table 1.14. In the first two columns, we consider the full sample of students. In the subsequent specifications, we include only students whose religious affiliation was.

The coefficient on the indicator variable for a college being affiliated with the same denomination as the student is consistently positive and robust across different specifications.This indicates that a college having the same denominational affiliation as the student is associated with an increase in the probability of a student choosing the college, conditional on other attributes.

Although omitted from the table, other coefficients retain the expected signs: students are more likely to choose colleges that are closer to their homes, charge less tuition, and are of higher quality (we proxy for institutional quality based on establishment date, total faculty, and 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 the marginal price or tuition.[21]

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 or her 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 corresponding feature.

For our sample of students, we find that in terms of both WTP and WTT, school denomination was a valued amenity. Based on conservative estimates from the full-sample, students on average have a marginal WTP of $77.94-$94.43 per semester for attending a college with affiliation identical to their own. Analogously, students were willing to travel up to 239-249 kilometers further to attend a college with a 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 affiliations matching their personal denominations. Denominational affiliation was a significant consideration in determining college choice. This implies that there were substantial returns or gains on the part of colleges from establishing denominationally 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-denomination schools suggest high returns to religious differentiation on the part of the college. The effect of religious competition on the market structure of higher

---

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

education is at least partially explained by the availability of religion as a dimension of product differentiation.

1.8 Conclusion

In this paper, we examine the religious origins of higher education in the United States. Specifically, we argue that religious competition, primarily within Protestant denominations, expanded the market provision of human capital during the 19th century. We present empirical evidence based on a panel of US counties to support this idea. To generate exogenous variation in denominational diversity, we also exploit the circumstances surrounding the Second Great Awakening as a source of identification. Our paper's central findings suggest that an unintended and previously unexplored consequence of religious diversity in the United States was an increased investment in higher education.

There is broad consensus among policymakers and researchers alike that universities play a significant role in economic development. At a cross-country level, an exceptional aspect of the contemporary United States is the strength of its higher education system, in terms of both the quantity of schools and the quality of those institutions. While the results in this paper pertain only to the supply of private denominational colleges, we contend that this historical episode brought short-term and long-term benefits to the development of the US collegiate system even after schools secularized.

In the short run, denominational colleges provided significant training and human capital prior to the large-scale development of land-grant universities. While denominational colleges were generally small institutions in terms of size, their educational importance should not be understated. In 1885, private denominational colleges conferred approximately 90 percent of all bachelor's degrees earned in the country.[22] These modest and spatially dispersed colleges were the backbone of American tertiary education throughout the late 19th century.

In the long run, the competitive landscape of private colleges provided an impetus for schools to become more consumer-oriented, to differentiate curricula and to invest in quality. As demand for scientific education rose following the Second Industrial Revolution, the US higher education system was well-positioned to capitalize on the excess capacity that had been built in the prior century (Labaree, 2017). What initially began as a humble patchwork of small and academically undistinguished institutions transformed into a system of reputable research universities. As a result, the United States became the global leader in the world market for higher education during the 20th century. The denominational colleges we study laid the foundation for this process.[23]

The "knowledge" industry remains, to this day, a key feature of "American Exceptionalism". US universities dominate global rankings: its top private research universities accumulate considerable wealth, attract talented students and faculty from abroad, and set the world's highest academic standards. This productive system is in part a consequence of unique circumstances in the 19th century: the absence of state-sponsored religion and the proliferation of Christian denominations. The relationship between religion and education is rich and nuanced, and our findings contribute to the understanding of this complex relationship in a setting with strong market forces and a divided church.

---

[22] The percentage is calculated from table 39 of the Report of the Commissioner of education in 1885. Additionally, our structural estimates of school choice indicate that students in the 19th century had strong preferences to attend colleges denominationally affiliated with their own church and closer to their hometowns; therefore, the entry of private denominational colleges likely increased total student enrollment. This is also corroborated in Figure1.7, which shows that the growth of colleges tracked the increase in student enrollment.

[23] Table 1.16 compares colleges built between 1850 and 1900 to colleges built after 1900, conditional on their survival. Colleges built in the latter half of the nineteenth century are significantly more selective in 2010.

Tables & Figures

Table 1.1: Descriptive Statistics: Denomination Market Share

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 1.2: 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 1.3: 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 1.4: Effect of Religious Rivalry on College Founding: Panel Estimates

Dependent Variable:	Ln(Colleges)				
	(1)	(2)	(3)	(4)	(5)
DenomFraction	0.3228*** (0.1188)	0.4865*** (0.1390)	0.3812*** (0.1368)	0.3662*** (0.1354)	0.3219** (0.1367)
County fixed effect	Yes	Yes	Yes	Yes	Yes
Decade fixed effects	Yes	Yes	Yes	Yes	Yes
Denomination Controls	No	Yes	Yes	Yes	Yes
Farm Productivity	No	No	Yes	Yes	Yes
Demographics Control	No	No	No	Yes	Yes
Manufacturing & RR	No	No	No	No	Yes
Observations	6487	6487	6487	6487	6487
R^2	.68	.69	.7	.7	.7
Oster's δ			2.384		

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 1.5: Testing Effect of Religious Rivalry on the Founding of Other Institutions

Dependent Variable:	Land Grant1890	State College1890	Public School1850	Public College1890
DenomFraction	0.0264 (0.0169)	0.0092 (0.0098)	-1.3024*** (0.3274)	0.0661 (0.0532)
County level controls	Yes	Yes	Yes	Yes
Denominational fixed effects	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes
Observations	2,634	2,634	1,600	2,600
Clusters	49	49	36	49
R^2	0.096	0.062	.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 1.6: Robustness: Excluding Regions

Dependent Variable:	Ln(Colleges)				
Sample:	All States (1)	Exc. South (2)	Exc. Midwest (3)	Exc. West (4)	Exc. Northeast (5)
DenomFraction	0.3219**	0.2197	0.4499***	0.2712*	0.2599*
	(0.1367)	(0.1750)	(0.1664)	(0.1383)	(0.1385)
Observations	6,487	4,068	4,687	6,006	5,705
R^2	.7	.7	.7	.71	.68

Notes: *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 1.7: Effect of Religious Rivalry on College Founding: State-Level Estimates

Dependent Variable:	Ln(Colleges)			
	(1)	(2)	(3)	(4)
DenomFraction	1.1171***	1.0480***	1.1022***	1.1442***
	(0.2841)	(0.2629)	(0.3365)	(0.3136)
State fixed effect	Yes	Yes	Yes	Yes
Decadal fixed effect	No	Yes	Yes	Yes
State-year trends	No	No	Yes	No
Time-varying controls	No	No	No	Yes
Observations	142	142	142	142
R^2	.91	.93	.95	.93

Notes:

Table 1.8: Effect of Religious Rivalry on College Founding: Alternative Treatment Def.

Dependent Variable:	Ln(Colleges)					
	(1)	(2)	(3)	(4)	(5)	(6)
# of Denominations	0.0782***	0.0495***	0.0384**	0.0309**	0.0269*	0.0269*
	(0.0134)	(0.0156)	(0.0154)	(0.0155)	(0.0156)	(0.0156)
County fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Decadal 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 Controls	No	No	No	Yes	Yes	Yes
Manufacturing & RR	No	No	No	No	Yes	Yes
Observations	6487	6487	6487	6487	6487	6487
R^2	.69	.69	.7	.7	.7	.7

Notes: The table reports log-linear estimates where the unit of observation is a county. $DenomFraction$ is the number of denomination present in a county-year. 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 1.9: Descriptive Statistics: New York State Towns

Variable	Revival towns		Non revival towns	
	mean	sd	mean	sd
Total number of revivals	4.769	4.22	0.000	0.00
Total number of revivals in 1831	1.700	2.00	0.000	0.00
Population 1820	2607.312	2193.03	1624.286	1520.99
Population 1840	4506.841	4013.82	2342.567	2027.38
Manufacturing share of population	0.037	0.02	0.033	0.02
Agriculture share of population	0.209	0.07	0.199	0.06
Commerce share of population	0.004	0.01	0.003	0.01
Total Area	85.763	88.68	62.859	107.49
Distance to waterways in 1831	30.645	27.84	26.535	23.13
Mean elevation	976.026	517.75	1020.633	656.74

Notes: The table reports the mean and standard deviation of variables for NY towns with at least one revival meeting and NY towns without any revivals. Population, manufacturing share of population, agriculture share of population and commerce share of population were collected from the New York censuses of 1820 & 1840. New York City and Albany are excluded.

Table 1.10: Placebo Test of Instrumental Variable on Various Outcomes

Dependent Variables:	(1) College Pre 1830	(2) Population 1840	(3) Manufacture 1840	(4) Commerce 1840	(5) Agriculture 1840
Dist. to Rochester	-3.22e-05	-0.0001	0.0003	0.0014	-0.0003
	(3.89e-05)	(0.0003)	(0.0004)	(0.0007)	(0.0003)
Population composition	Yes				
Township level Controls	Yes	Yes	Yes	Yes	Yes
Geological Controls	Yes	Yes	Yes	Yes	Yes
Lag outcome	Yes	Yes	Yes	Yes	Yes
Observations	521	521	521	521	521
R^2	0.026	.72	.54	.33	.54

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 1.11: Revivalism and colleges: First Stage & Reduced Form

Dependent Variable:	Colleges	Revivals	DenomFraction
Dist. to Rochester	-0.0002**	-0.0100***	-0.0003***
	(0.0001)	(0.0016)	(0.0001)
Geological Controls	Yes	Yes	Yes
Township Level Controls	Yes	Yes	Yes
Observations	522	522	504
R^2	0.086	.31	.17

Notes: The table reports linear regressions. The dependent variable in column1 is the number of Colleges built between 1830 and 1860. The dependent variable in column2 is the number of revivals between 1825 and 1835. The dependent variable in column3 is the constructed $DenomFraction$ in 1845.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 is excluded from both samples. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 1.12: Revivalism and colleges: Second Stage

Dependent Variable: Colleges	2SLS	2SLS	2SLS	OLS	OLS	OLS
DenomFraction1845	0.5413***	0.5045***	0.4979**	0.0261	0.0299	0.0064
	(0.1865)	(0.1655)	(0.1993)	(0.0199)	(0.0224)	(0.0242)
Revivals in 1831			-0.0052			0.0025
			(0.0057)			(0.0051)
Revivals before 1831			0.0052			0.0118*
			(0.0068)			(0.0064)
Geographic Controls	No	Yes	Yes	No	Yes	Yes
Town Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	504	504	504	504	504	504
R^2				0.06	0.07	0.10
$F-Stat$	15.921	19.013	9.554			

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. Total revivals equals the total number of revival meetings from 1825-1835. 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 is excluded from both samples. Kleibergen-Paap F-statistic is reported. Standard errors are clustered at the 1820 county level. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 1.13: Student Micro-Data Summary Statistics

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: Panel A gives descriptive statistics for our student-college linked data. We report separately mean and standard deviation of students whose religious affiliation was available and students whose religious information was missing. Panel B reports summary statistics of colleges attended by all students in our sample.

Table 1.14: College Choice, Conditional Logit Coefficients

	Parameter Estimates			
	Full Sample		Relig. Sample	
	(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
College county fixed effect	No	Yes	No	Yes
Observations	258768	250600	114552	110810

This table shows results from estimating the conditional Logit model discussed in Section 1.7, 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 1.15: Short Run College Quality

Dependent Variable:	Δ Students	Δ 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	232	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 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.

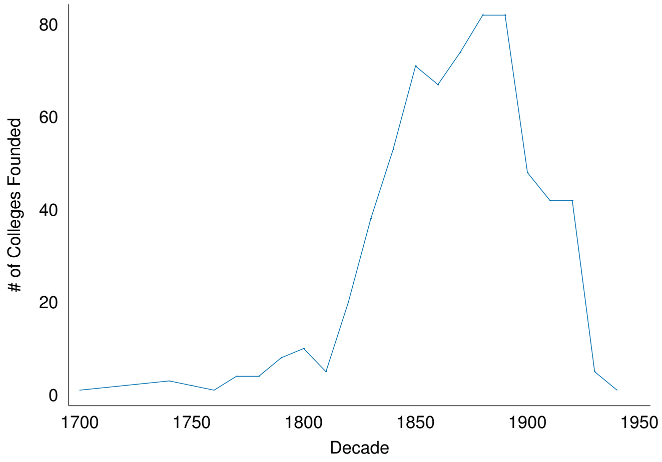
Table 1.16: Long Run College Quality

Dependent Variable:	%Women Admitted	%Men Admitted	SAT verbal	SAT math	Carnegie Classification
Historical Colleges	-0.072***	-0.053***	51.29***	54.11***	8.6944***
	(0.0170)	(0.0172)	(5.3268)	(5.4157)	(0.3460)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2186	2180	1199	1214	7275
R^2	0.0540	0.0646	0.1691	0.1680	0.1064

Notes: The table reports OLS estimates. An observation is a college built after 1850 that was reported in the IPEDS 2010 survey. The dependent variables are shares of women admitted in column 1, shares of men admitted in column 2, SAT verbal score in column 3, SAT math score in column 4, and Carnegie Classification in column 5. Details of Carnegie Classification could be found here (`https://carnegieclassifications.iu.edu/classification_descriptions/undergraduate_profile.php`). Full-time four-year, more selective, higher transfer-in receives the highest Carnegie score of 13. Variable of interest, historical colleges, equals 1 if a college was built between 1850 to 1900.

*** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

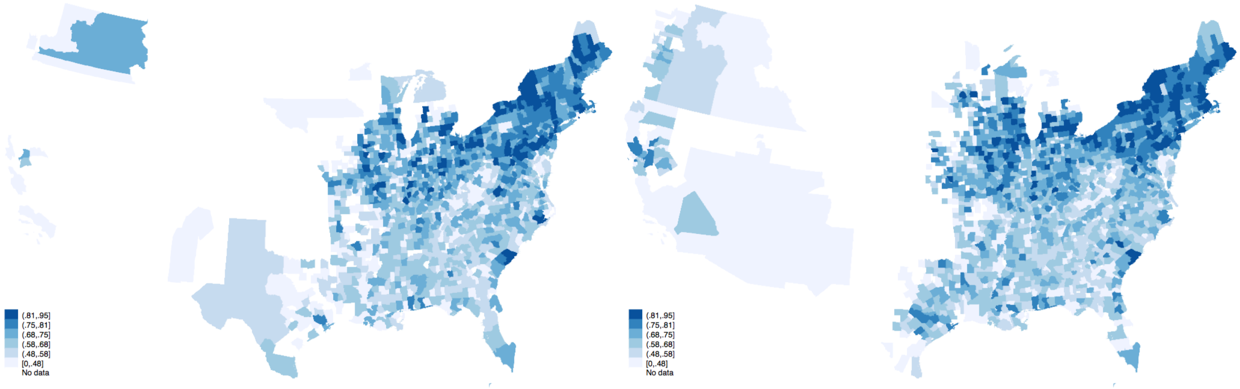
Figure 1.1: College Expansion



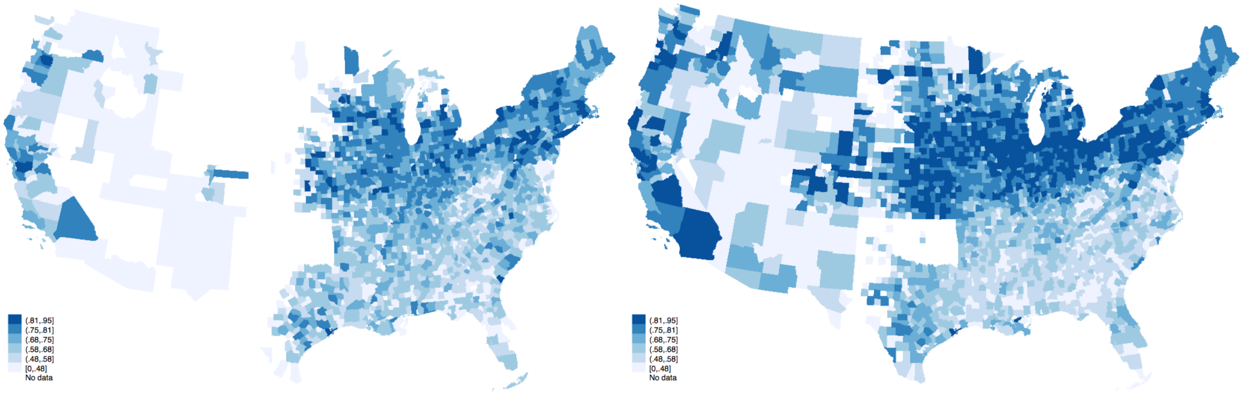
Notes: The figure reports the number of surviving colleges built in each decade from the 1700s to the 1930s. Source: Goldin and Katz (1999)

Figure 1.2: Denominational Competition: 1850-1890

(a) 1850 (b) 1860



(c) 1870 (d) 1890



Notes: Figures a, b, c, and d map $DenomFraction$ for each county in the corresponding decade. $DenomFraction$ 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.

Figure 1.3: College Expansion by Institutional Type

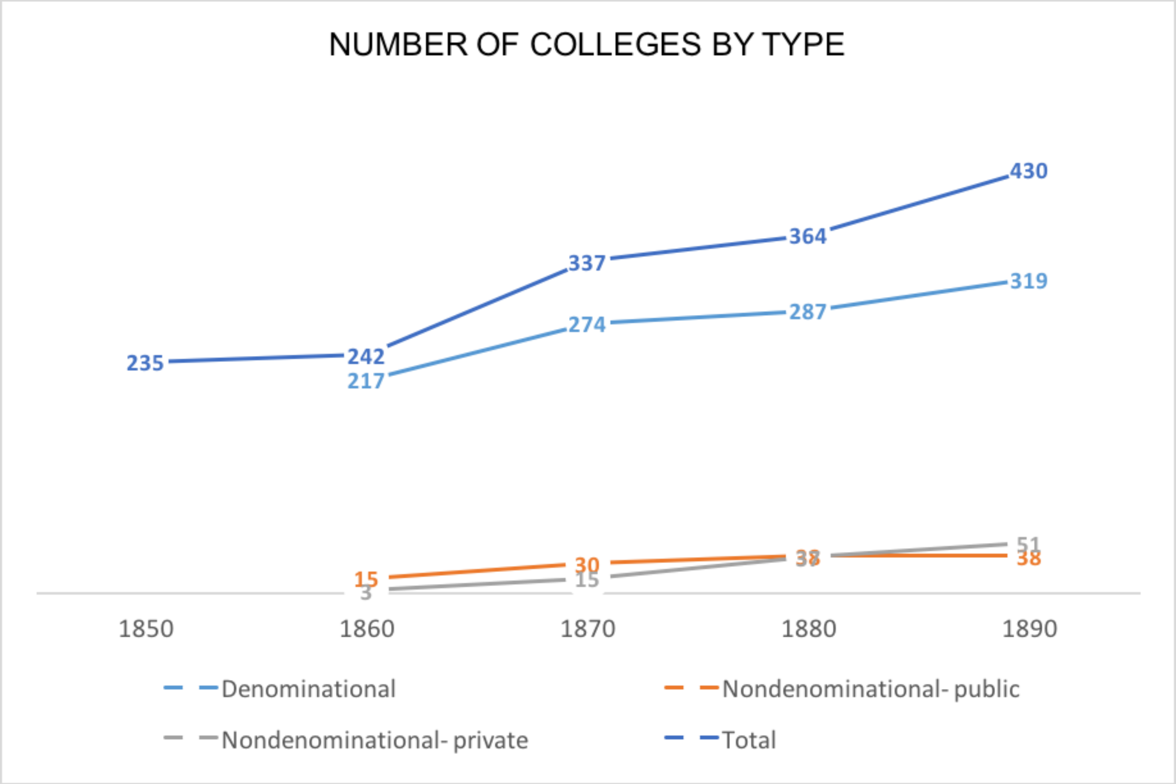
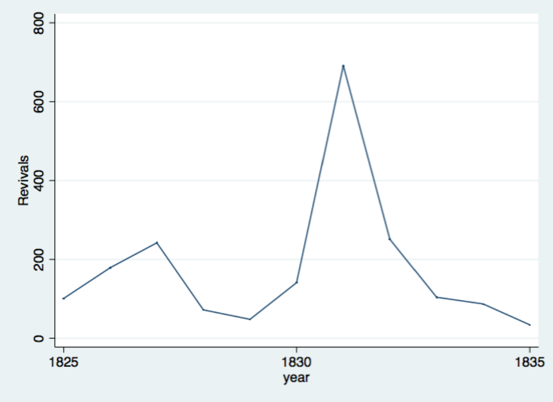


Figure 1.4: Secularization of higher education over time

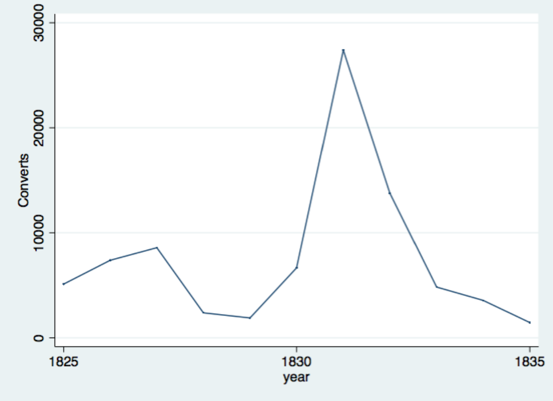
Notes:

Figure 1.5: 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

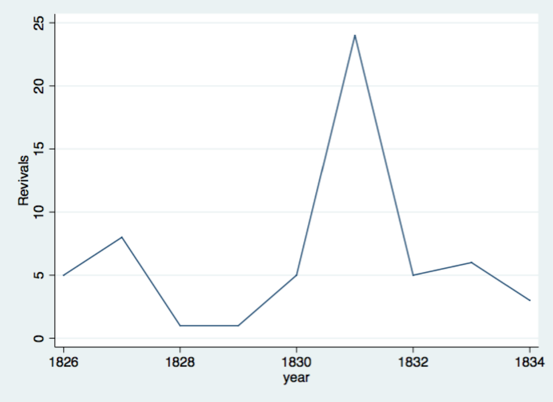
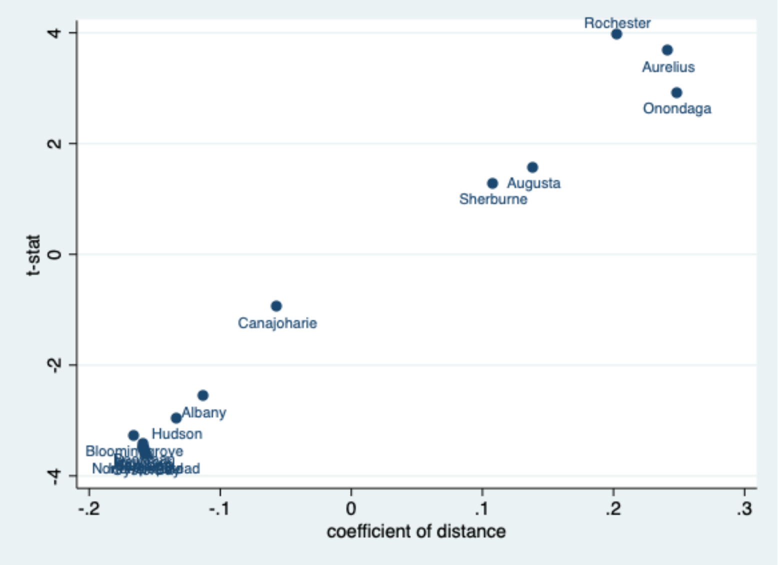


Figure (a), Figure (b) and Figure (c) present the trend of revival activity measured by total number of revival meetings, total number of converts and total number of affected counties respectively.

Figure 1.6: Placebo First Stage

(a) Distance Coefficient on Converts



(b) Distance Coefficient on Revivals

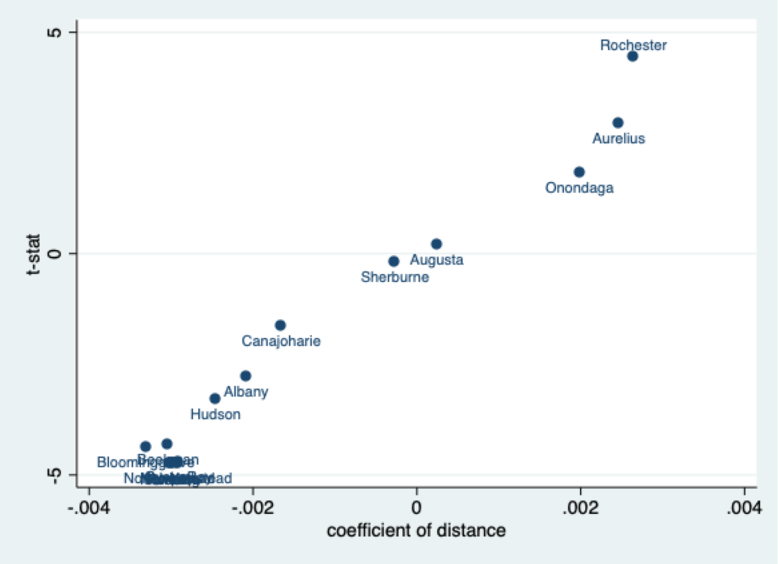
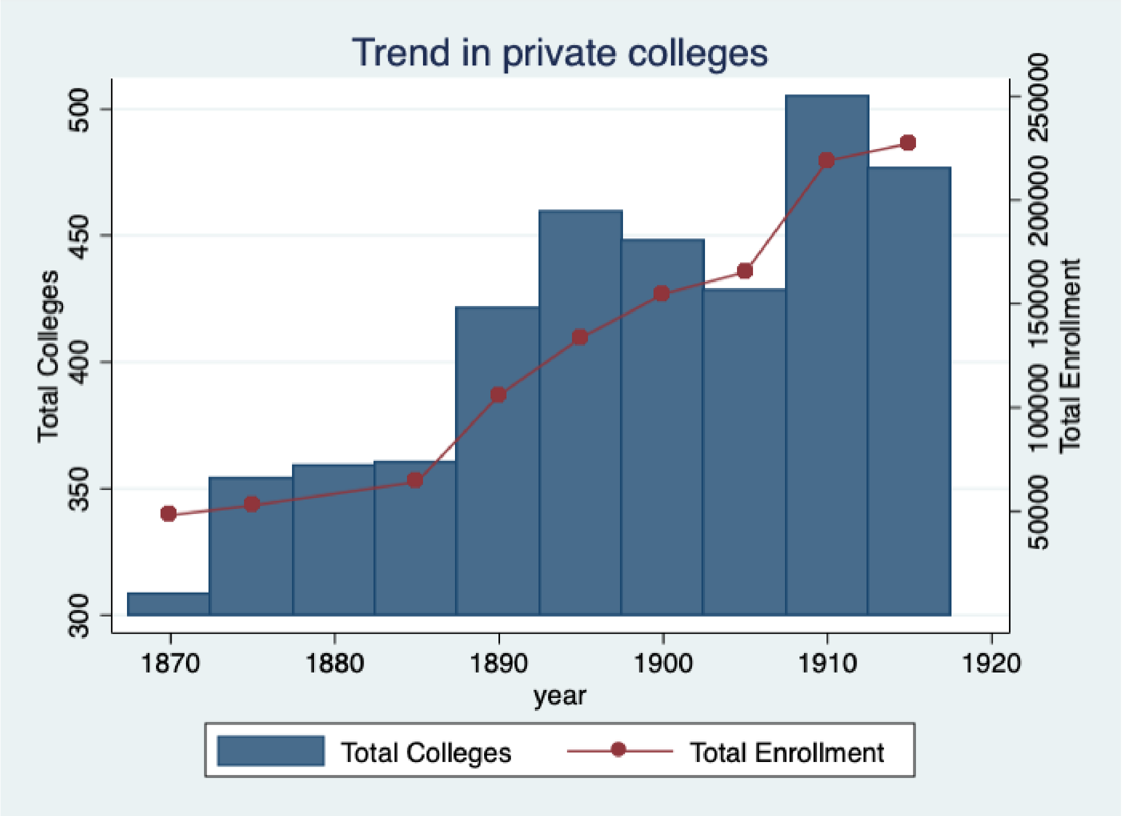


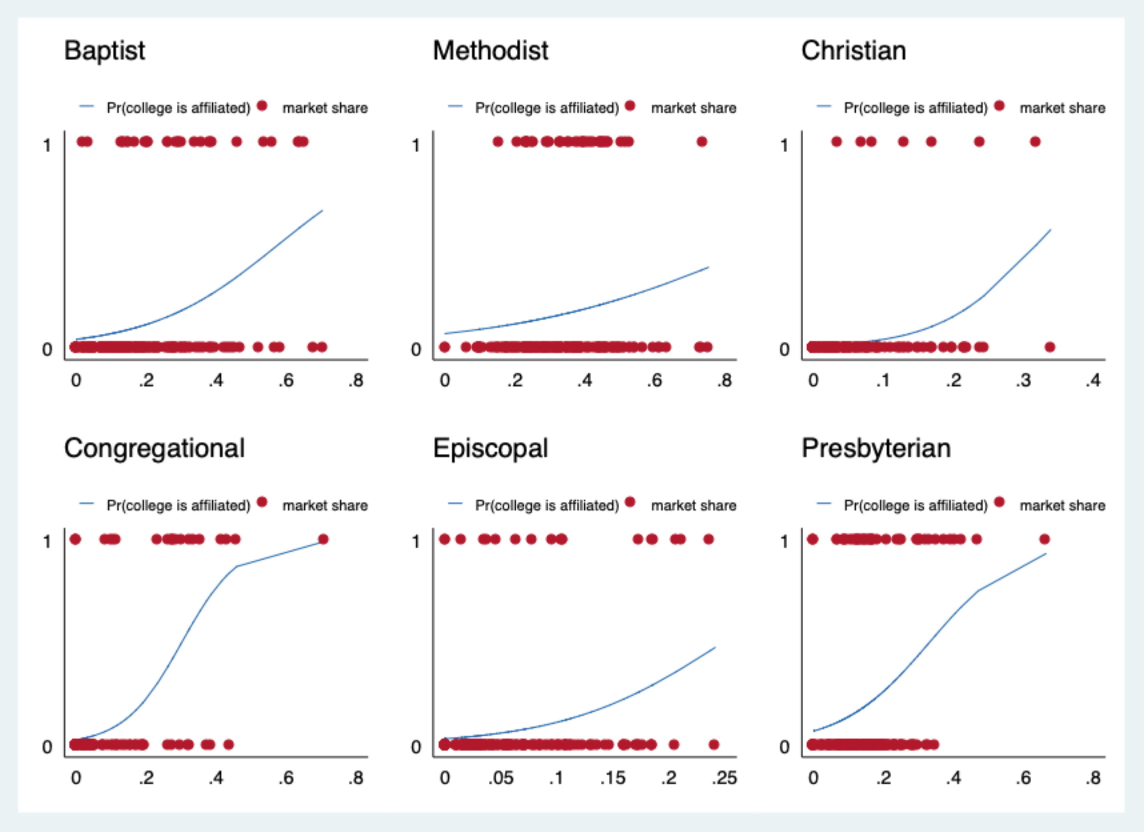
Figure a and Figure b draw comparisons between the first stage coefficients from Columns 1&2 of Table 1.11, respectively, and the coefficients from the same specification, but estimated with the distance from 16 New York towns with population greater than 10,000 by 1820, as opposed to Rochester. Dots represent the negative of first stage coefficients for 16 New York towns and Rochester

Figure 1.7: Growth of total students in private colleges



Data Source: Report of the Commissioner of Education

Figure 1.8: Denomination Market share & affiliated college



Notes: each figure shows the predicted probability that the college is affiliated with a certain denomination based on the county-level market share of that denomination. Data are from the 1860 sample.

CHAPTER 2

HOME ECONOMICS AND WOMEN’S GATEWAY TO SCIENCE

2.1 Introduction

In the past century, the gender gap in overall college enrollment and enrollment in each STEM major (with one notable exception of computer science) has reversed or narrowed (Goldin, 2014; Goldin and Kuziemko, 2006). However, the speed of convergence varied very much across fields (Kahn and Ginther, 2017). Within STEM, the share of women in life sciences and chemistry grew faster than other physical sciences and engineering. This pattern of preferences has been documented for women from different samples and periods (Lubinski and Benbow, 1992; Xie and Shauman, 2003). Because physical sciences and engineering are more math-intensive, the existing literature has been primarily concerned with gendered differences in math ability and whether the differences are caused by nature or culture. Kahn and Ginther (2017) thoroughly surveyed the literature on women’s under-representation in the math-intensive fields. So far, scholars seem to take for granted that women are well-represented in life sciences. The literature has vaguely attributed it to preferences (Wiswall and Zafar, 2015) and psychological explanations (Ceci et al., 2014) without addressing where the preferences or mentality come from.

Our paper explores the higher education history in the U.S. and seeks the origins of female preferences for life sciences. Specifically, we look at collegiate Home Economics, a women’s field developed around the Progressive Era. Home Ec was an academic entrepreneurship that professionalized different aspects of homemaking. It did so by incorporating science courses, especially chemistry and biological sciences, into its core curriculum. We document the science courses that frequented the Home Ec curriculum and establish a short-run relationship between Home Ec and women’s enrollment in science. Given that gendered attitude, particularly in academic stereotypes, may be transmitted across generations (Alesina et al., 2013; Eccles and Jacobs, 1986), the development of women's curricula in the early twentieth century may have long term influence on observed academic preferences today.

For most of the nineteenth century, private colleges and classical curricula dominated the higher education landscape. Men and women had a homogeneous studying road map, and the aggregate gender ratio in classics was balanced.[1] Their paths diverged when public institutions and practical education rose to prominence during the second industrial revolution. Since the practical fields correlated closely to specific occupations, so did the corresponding gender ratios. For example, agriculture and engineering were almost exclusively male now that farmers and mechanics were primarily men. The founding of Home Ec followed the gender norm that most women became homemakers and thus aligned women's practical education with their primary vocation. Because this view was widely accepted 100 years ago, Home Ec was conveniently legitimized at land-grant universities and received generous public funding. The founding of Home Ec presented an extreme example where a woman's human capital investment was a backward induction from her expected role in married life. However, from Becker (1985)'s discussion in the sexual division of labor to Bursztyn et al. (2017) 's experiment on marriage market signaling, the literature still reminds us of the relevance of traditional gender roles in women's education choices.

Despite the conservatism in its appearance, the founders of Home Ec were ambitious and progressive. They were concerned with expanding academic and professional avenues for women. Students who were interested in homemaking complained that the curriculum was not practical. From the founders' point of view, a rigorous scientific standard was a necessary foundation for new career paths, since "they cannot all become teachers".[2] The founders forged careers for Home Ec graduates in hospitals, food and consumer goods industries, and the public sector of food testing and regulation (Nyhart, 1997). The availability of applied science jobs for women may loop back to a higher interest in them studying science.

---

[1] According to the Report of the commissioner of education in 1910, 34,492 women and 36,077 men enrolled in classics.

[2] Quoted by Ellen Richards in (McCollough, 1912)

To test the hypothetical link from Home Ec to science, we collected college level data from the Report of the Commissioner of Education in 1910, which recorded enrollment by major and gender. For each institution, we also document the land-grant status, endowment by private and public source, value of scientific apparatus, teaching disbursement, etc. The first part of our empirical analysis reveals a positive significant relationship between Home Ec and women's science enrollment in cross section. The spill-over rate from Home Ec to science is approximately 10:1. The estimated spill-over rate is robust to controlling for various sets of confounding factors.

A key concern with the cross-section analysis is that colleges with Home Ec programs vary in unobserved characteristics, such as their acceptance of highly educated women. In that case, colleges with larger Home Ec programs had more women in all majors. To address this, we conduct falsification checks. Placebo analysis shows that no other major that women were likely to choose had a positive relationship with women in science. Home Ec had singular importance in driving up women's enrollment in science. To exclude the possibility that colleges with Home Ec had a higher demand for science, we show that Home Ec does not predict men's enrollment in science.

To identify the causal effect of Home Ec, we adopt an instrumental strategy that exploits the institutional setup of Home Ec at land-grant universities: it was coded as an agricultural subject. From the start, Home Ec programs were closely tied to agricultural colleges, sharing administration, courses, and resources. For instance, Home Ec was entitled to funding from the Hatch Act of 1887 and the Smith-Lever Act of 1914, both agricultural grants from the federal government. Building on this insight, we instrument for Home Ec using the size of enrollment in agriculture. We confirm that the land-grant with higher agricultural demand had larger Home Ec programs, translating into more women in science.

To study Home Ec's short-run effect on narrowing the gender gap in science, we compiled a panel dataset of 21 colleges spanning from 1910 to 1940. The panel dataset is assembled from college yearbooks, which typically contain full student names and declared majors. We infer a student's gender by matching the first name to the 100% U.S decennial population census in the closest decade and compute the probability that the first name is male. We then aggregate the gender ratio for science majors for each college-year. Using a fixed effect model, we show that the gender gap in science was smaller in the presence of Home Ec. The gender gap in science also shrank as the proportion of women taking Home Ec increased.

The rest of the paper is organized as follows: Section 2.2 provides a brief history on the founding of Home Ec. Section 2.3 explains the sources of our datasets. Section 2.4 describes our empirical specifications for cross section, describes the instrument, and shows the results. Section 2.5 presents the panel analyses on the opening of new Home Ec program, utilizing a novel dataset in student yearbooks. Section 2.6 describes potential mechanisms. Section 2.7 concludes.

2.2 Historical Background

The history of Home Ec education intertwined with that of the land-grant universities. Land-grant universities are higher education institutions tasked with teaching agriculture, military tactics, and the mechanical arts as well as classical studies to that members of the working classes by the Morrill Acts of 1862. [3] This mission contrasted with the historical practice of a liberal arts curriculum provided by private institutions.

Home Ec developed as a feminine parallel to agricultural education, a key component among the land-grant missions. Agriculture studies rarely existed among other state universities and were the primary target for federal financial support. The Hatch Act of 1887 provided federal funds to states to establish a series of agricultural experiment stations under the direction of each state's land grant college. The act also granted annual appropriations for research on the condition that state funds matched those funds. Besides agricultural research, the experiment stations were responsible for extension work: bringing new information and the results of agricultural research into rural areas. The outreach mission was further expanded by the Smith-Lever Act of 1914. Andrews (2019) discusses the causal effect of land grant colleges on local agricultural output.

Since agriculture and mechanics were supposedly practical fields for men, only a small number of women attended land grant universities (25.1% of student population in 1910) in comparison

---

[3] Land-grant universities were designated by states to receive the benefits of the Morrill Acts of 1862 and 1890.The first Morrill Act granted federally controlled land, hence land-grant, to the states to sell and raise funds.

to private colleges (47.8%) or other state universities (34.3%).[4] Some advocated to enroll more women in land-grant universities. For example, trustees at Iowa State College declared in 1869, “if young men are to be educated to fit them for successful, intelligent and practical farmers and mechanics, is it not as essential that young women should be educated in a manner that will qualify them to properly understand and discharge their duties as wives of farmers and mechanics? (Epright and Ferguson, 1971)” In order to increase women’s enrollment at land-grant universities, they created a practical field for women, Home Ec.

The U.S. Department of Agriculture (USDA) was a significant impetus to the development of Home Ec. Through coding Home Ec as a subject under the broad umbrella of agriculture, it enabled Home Ec related research and outreach programs to unlock funding from both the Hatch Act and the Smith-Lever Act. As we will demonstrate, federal financial support significantly increased program sizes of Home Ec. Besides providing funding to Home Ec programs, the USDA provided employment opportunities for Home Ec graduates. It opened the Office of Home Economics in 1915, where home economists were hired to work with agricultural experts (Elias, 2008).

Home Ec’s association with agriculture also influenced its curriculum. Many Home Ec departments started as a part of agricultural colleges and adopted the core courses from agriculture programs.[5] Since agriculture is a multidisciplinary field of biology, students who enrolled in Home Ec programs were fully exposed to the foundation of biological sciences. Before taking specialized courses in Home Ec, they have typically completed courses in chemistry, physiology, and bacteriology.

While agriculture played an important role, the Home Ec curriculum was equally a product of its time. The late nineteenth century and early twentieth century witnessed scientific discoveries and culture shifts that directly shaped homemaking (Mokyr, 2000). Home Ec began to take shape as an academic field around 1900, when the first Home Ec subject, Hygiene and Sanitation, emerged (Elias, 2008). The timing coincides with the breakthrough in bacteriology. Scientists

---

[4] The percentages are derived from Table 16 of the Report of the commissioner of education in 1910.

[5] For example, when Iowa State University inaugurated the Home Ec program, the freshmen year courses for Home Ec were identical with that in agriculture. The sophomore year courses continued to include courses in chemistry and botany (Eppright and Ferguson, 1971).

came to understand that germs caused disease and infection, and they are preventable through hygiene and sanitation prevented them. The second Home Ec subject, Food and Nutrition, formed in the late 1910s and integrally linked to the discovery of vitamins and minerals. In the 1930s, Home Ec training added another component in developmental psychology, possibly in response to the cultural shift in the belief that children deserved protection, nurture, and education.

The Home Ec movement was most relevant in the first half of the twentieth century. The number of degrees conferred in Home Ec grew substantially until the 1950s (see figure 2.2).[6] This period concurred with women's increasing productivity in the labor market and at home. This was when high school attendance increased significantly (Claudia Goldin, 1999), and electrification brought household labor-saving technologies (Cowan, 1985) and skill-biased job opportunities (Gray, 2013; Vidart, 2020) in the service sector (Goldin, 2006). However, women's increasing labor supply did not correlate with a decline in the time spent on housework (Cowan, 1985; Ramey and Francis, 2009). Mokyr (2000) reconciled the paradox by accounting for the growing domestic knowledge of hygiene, nutrition, and child development. The information was transmitted to households via home economics, more specifically, women who became high school teachers and extension workers in home economics (Stage and Vincenti, 1997).

2.3 Data

The *Report of the Commissioner of Education* provided data on enrollment by major and sex at the college level in 1910. The information was self-reported. A total of 583 institutions completed the survey, and nine majors were classified: classical & general culture, general science, agriculture, household economy (Home Ec), engineering, education, commerce, music, and fine art. Figure 2.1 shows the distribution of majors at different types of institutions. Classical education dominated the higher education landscape at private colleges for both men and women. At land-grant universities, roughly 20% of men enrolled in agriculture, and 20% of women enrolled in home economics. While agriculture was supposedly a men's field and home economics, a women's, there were a

---

[6] The ratio of conferred degrees on enrollment is approximately 1:10 for women, based on 1910 statistics.

few exceptions. Women constituted 1.7% of agricultural students in 1910, and although home economics was exclusively female in 1910, some men studied home economics in later years.[7]

General Science was not a popular major for either gender. However, the share of women in general science (women in general science/total female students) was slightly larger at land-grant colleges (5.4%). The percentage was 1.5% and 3.8% at state colleges and private colleges, respectively. Men were more likely to major in science in private college (11.4% of men) compare to state (3.5%) or land-grant (4.3%) colleges.

Besides enrollment, the commissioners collected information on many college characteristics, including location, the number of faculty, founding date, funding sources, values of various assets, library volumes, tuition costs. For land-grant universities, in particular, teaching expenses on different subjects were reported.

Unfortunately, the commissioners only collected data on Home Ec and science degrees jointly in 1910, and the series of *Annual Report of the Commissioner of Education* ended in 1915. The follow-up series of *Biennial Survey of Education* collected data on enrollment for professional majors and arts & sciences majors, but not on science majors separately. Therefore, we cannot observe changes in the women's science enrollment over time at a national scale.

To overcome this challenge, we resort to the historical student yearbooks that are available through ancestry.com. The college yearbooks include full student names and hometown, which can be used to infer gender and to match students from yearbooks to other data sources such as the patent record. Instead of categorical disciplines, we observe enrollment in specific majors and double majors, if any.

We collected yearbooks from 21 different land-grant colleges, covering 305 yearbooks from 1879 to 1940 and including records for 83,448 undergraduate seniors. Summary statistics of all

---

[7] Elias (2008) suggested that when men enrolled in home economics courses, these were most often institutional management classes, a group of topics that later became the core of hotel management. Using Cornell University's yearbooks from 1919 to 1936, we estimate that 95 percent of male students in Home Ec indeed majored in Hotel Management. However, data from the 1958 *Biennial Survey of Education* did not show a specific concentration. Among the 36 bachelor's degrees in home economics conferred to men, there were 2 in general curriculums, 2 in child development and family relations, 3 in clothing and textile, 11 in foods and nutrition, 6 in institution management, and 12 in other unspecified home economics fields.

colleges appearing in the sample are provided in Table 2.7. In our sample, 7 colleges had no Home Ec enrollments at all, 5 colleges had positive Home Ec enrollment for all transcribed years, 9 colleges went from no enrollment to positive enrollment in Home Ec. Figure 2.3 (a) (b) show the increase in the number of Home Ec programs and total Home Ec enrollment.

We use first names from the US decennial censuses to infer the gender of each student. For each state and each census, we calculate the probability of being male for each first name, and then impute the inferred gender of the student. Similar technique has been used in (Andrews, 2019) to infer gender and race of patentees, (Cook et al., 2014) to identify race, and (Jones, 2009) to infer age.

We grouped home economics majors and science majors to the best of our ability to account for differences in the granularity of major names and formats. We excluded majors in Arts & Sciences since we cannot identify whether they are science majors or not. For this reason, the enrollment in sciences is under-counted, and more so in recent years than earlier years due to the increasing number of Arts & Sciences major reported. Therefore, we focus on gender ratio in science, an alternative measure to the number of women in science that is less sensitive to under-counting. Figure 2.6 shows the gender ratio in science over time for three college groups: colleges that never had Home Ec enrollment (the "never" group), colleges that had Home Ec enrollment throughout the years (the "always" group), and colleges that switched from zero to positive Home Ec enrollment (the switching group). The fraction of women in science is consistently higher in the "always" group than colleges in the "never" group. The trend in the switching group follows the "never" group in earlier years and converges to the "always" group in recent years.

We show the share of women in science (number of women in science/ total women) for a cross-sectional comparison in figure 2.4. Even though this measure loses precision over time, it offers a more direct interpretation of women's participation in science. As we have expected, the share of women in science is consistently higher in colleges with Home Ec than colleges without Home Ec.

2.4 Cross Section Analysis

This section presents the empirical exploration of the relationship between Home Ec and women's in science through OLS estimation. A discussion of endogeneity concerns and corresponding instrumental variable solution is presented in the next section. We focus on land-grant universities because of their comparable institutional setup and academic standards for Home Ec programs. We include the regression for the full sample in the appendix.

2.4.1 Land Grant Sample

The cross section estimating equation is:

$$number\ of\ women\ in\ science_c = \beta \cdot number\ of\ women\ in\ home\ ec_c + \theta X_c + \epsilon_c \tag{2.1}$$

where the outcome is the number of degrees in general science conferred to women in 1910 at college c and the variable of interest is the number of degrees in home economics conferred to women in 1910 at college c. X_c is a vector of control variables, and ϵ_c is an error term.

The coefficient of interest, β, captures the relationship between home economics and women in science. A positive β provides evidence that women who enrolled in home economics courses spilled over into science majors.

We consider four subcategories of controls to capture the differences across colleges that might affect the number of women in science: the size and breadth of academic offerings, the types of funding received, the overall resources available, and the emphasis on science education. The baseline controls consist of the number of women graduating in classics, the number of majors available, and the size of the student body. Column 1 in table 2.1 shows this baseline estimate. As predicted, the number of women in home ec strongly predicts that in general science: an increase of 10 women in home ec is correlated with 1 more woman in general science. Since classics was still the most popular major for women at land-grant, the number of women in classics serves as a natural comparison group. In contrast to home ec, classics show no signs of spill-over to women in general science.

Even though land grant universities provided similar education to each other, they still could have varied in their commitment to science depending on the geographic location or the attitude of the university administration. Therefore, we quantify the supply and demand for science education. We control the sunk investment in scientific instruction by the value of scientific apparatus and machinery as of 1910 and the marginal spending on science education by the disbursement of funds towards the teaching of natural sciences in 1910. Additionally, we control for the overall demand for science education through the number of men in general science. Column 2 in table 2.1 show estimates after controlling for the popularity of the science major in different ways. The controls, especially the number of men in science, positively correlate with the number of women in science. Reassuringly, they hardly affect the coefficient of interest.

We further test robustness on university resources to account for the overall quality of education: the number of faculty, the volume of library books, and total endowment. Column 3 in table 2.1 shows that estimates are robust to controlling for the university resources.

Controls on funding include funds received from government sources and funds collected from private sources. Funds from government sources, in particular the Hatch Act funds, were directed to agricultural, home economics, and biological research. The spill-over effect may be caused directly via the Hatch Act spending. Table 2.2 shows the direct effects of public and private funding on various subjects. While public financing positively predicts the size of agriculture and home economics programs, it has no direct impact on the size of the science program. Consistent with this result, column 4 in table 2.1 shows that estimates are robust to controlling for types of funding.

2.4.2 Robustness Checks

The size of a Home Ec program may have indicated the degree to which women receive higher education was accepted. To account for this possibility, we replicate our regression for all other majors which women were likely to take. Columns 1-6 of table 2.3 show that no major other than

Home Ec had a positive and significant impact on women's enrollment in science. This evidence supports the singular importance of Home Ec in driving women to science.

To further demonstrate that the relevance of Home Ec is not driven by demand for science in general, we run a placebo regression on men in science. We show in Column 7 of table 2.3 that no significant relationship exists between Home Ec and men's enrollment in science.

2.4.3 Agriculture Instrument

What drives the underlying cross-sectional variation in Home Ec program sizes? Understanding the driving force will allow us to interpret whether our estimates are causal or merely correlational. Even though all land-grant universities were encouraged to establish the Home Ec program, not every institution opened a Home Ec major, and sizes of the program varied significantly.

As discussed in the historical background, the U.S. Department of Agriculture contributed to Home Ec development by including it as an agricultural subject. Thus, the in-state demand for agricultural education likely correlated with the offering for Home Ec. Specifically, a higher in-state demand for agricultural education led to a larger agriculture program; since Home Ec shared the same funding sources with agriculture, a larger agriculture program meant a lower marginal cost to invest in a Home Ec program. Home Ec programs were often part of schools of agriculture, Home Ec research were conducted at the Agricultural Experiment Stations, and collaborations between Home Ec faculty and agriculture faculty were common (Smith, 1933). Building off from this observation, we use the number of men in agriculture program as an instrument for the number of women in Home Ec.

Before turning to the 2SLS results, we will examine the first stage relationship in our land-grant sample. Columns 1&2 of table 2.4 show the coefficient on the instruments in predicting the Home Ec program size in 1910. The relevance assumption is satisfied as colleges with larger agricultural programs had larger Home Ec programs. Figure 2.8 visualizes the relationship. Unfortunately, 1910 was still early for some colleges to establish a Home Ec program. For instance, Cornell University's large Home Ec school that pioneered the Hotel Management program, didnt begin until 1919.[8] Hence, our instrument worked better at the intensive margin than the extensive margin. Because sizes of the agriculture program only explain a proportion of variation in sizes of the Home Ec program (when Home Ec program had been established), the F-stat is small and our 2SLS estimates would be biased towards OLS estimates given the small sample (Bound et al., 1995).

Obviously, states with high demand for agricultural education were not randomly chosen. The states with the largest agriculture program in 1910: New York, Pennsylvania, and Illinois, were among the states with the highest value of farm building and equipment, an indicator for the mechanization of agriculture.[9] The adoption of new farming technology likely incentivized potential farmers to attend college and learn about the latest science and techniques in agriculture. To confirm that farm mechanization is the primary channel to agriculture enrollment, we compare the relative predictive power of other major aspects of agriculture: employment, output, livestock, size of farms, and the abundance of farms. Since the set of agricultural variables available from the 1910 Census of Agriculture is quite large, we use the Least Absolute Shrinkage and Selection Operator (LASSO) technique (Tibshirani, 1996) to select the predictors in forwarding steps. Table B.2 in the Appendix shows the variable selection procedure. Indeed, the value of farm equipment stands out as the most correlated agriculture predictor, and its combination with state population and student population gives the optimal Lasso solution.

The credibility of our research design hinges on the assumption that demand for agriculture did not affect women in science directly or indirectly for reasons other than Home Ec. The exclusion restriction is challenged if the relevance of agricultural education, whose variation could be attributed to the adoption of agricultural technology, increases the demand for science education. To rule out this possibility, we regress agricultural program sizes on different science measures: the number of men in general science, teaching expenses on science, and the value of scientific

---

[8] Home economics began as a department in the College of Agriculture at Cornell. In 1919, Cornells trustees made it the School of Home Economics (Engst and Friedlander, 2014). From Cornell's yearbooks, we first observe seniors majoring in Home Ec in 1919.

[9] Figure 2.7 shows the relationship between the enrollment in agriculture at land-grant college and the value of farm machinery of that state.

apparatus. Table 2.5 presents the results and shows that no strong correlation existed. In contrast, we found a significant reduced-form relationship between sizes of the agricultural program and the number of women in general science (table 2.4 columns 3&4).

IV results:

Now we turn to quantify the marginal effect of Home Ec program size on womens enrollments in general science in the second stage. Table 2.6 presents the estimates in 2SLS (columns 1-2). The results are robust to controlling for university scale measures and science-related conditions. Our instrument proved to be relevant, and the second-stage results are larger than the cross-sectional estimates: the spill-over ratio from Home Ec to general science is approximately 10: 1 according to the OLS estimation and 6.5: 1 according to the 2SLS estimates.

The discrepancy between IV and OLS estimates point to the fact that Home Ec program sizes could be correlated with omitted variables that are negatively associated with womens enrollment in general science, resulting in a downward bias in the OLS estimates. This suggests that if anything there is a negative bias in the selection of Home Ec programs with respect to womens scientific pursuit. This accords with the narrative evidence which indicates that the design for Home Ec was motivated by ideas about traditional gender roles and targeted women from rural areas (Schwieder, 1986). The bias could occur, for instance, if prior to the college entrance, a woman (or her parents) who chose the Home Ec major had a lower level of interest in science than another woman who chose the major in music or classics.

The analysis in this section offers two distinct advantages. First, although Home Ec education was a signature of land-grant universities, the instrument explains the variation in sizes of the Home Ec program within the land-grant sample. Second, the empirical design allows us to assess the plausibility of the identification strategy more easily. Since agricultural education was almost exclusively male[10] and has specific motivated origins that are bound to geological conditions and factor prices,[11] it should have little organic connection with women in higher education other than

---

[10] 98.3% of agriculture degrees were awarded to men in 1910.

[11] For example, Manuelli and Seshadri (2014) attributed the variation in horse prices for the case of tractor adoption.

the arbitrary Home Ec-Agriculture bundle assigned by the USDA. Altogether, this allows us to make progress toward actual causal estimation of the effect of Home Ec education on bringing women into science majors.

2.5 Panel Analysis with Student yearbooks

In this section, we expand the time horizon from 1910 to 1940 and test the hypothesis in the panel framework. We compile measures in enrollment by major and gender at the college level by using student data from the yearbooks. This dataset was first deployed in (Andrews, 2019). We introduce time variation and exploit both the establishment of new Home Ec programs and the relative sizes of the program at different points in time. In contrast to the previous analysis, here our estimates are identified not by variation across colleges, but variation across time in a given college. Due to the under-counting problem discussed in the data section, we modify our specification as follows:

$$frac.\ of\ women\ in\ science_{c,t} = \beta \cdot home\ ec\ dummy_{c,t} + \delta_c + \theta X_{c,t} + \epsilon_{c,t} \tag{2.2}$$

$$frac.\ of\ women\ in\ science_{c,t} = \beta \cdot share\ of\ women\ in\ home\ ec_{c,t} + \delta_c + \theta X_{c,t} + \epsilon_{c,t} \tag{2.3}$$

where outcome of is the fraction of women in science (number of women in science/total science enrollment), $home\ ec\ dummy_{c,t}$ is a dummy variable that equals 1 if Home Ec had positive enrollment at college c in year t, $share\ of\ women\ in\ home\ ec_{c,t}$ is the share of women in Home Ec (number of women in Home Ec/ total female students) at college c in year t. The inclusion of college fixed effect, δ_c, allows us to flexibly address any time invariant characteristics which may differentially affect demand for practical or scientific education. Factors may include geological conditions, distance to metropolitan areas, founding philosophy of a institution, etc.

Since our sample is small and unbalanced, we control for year intervals instead of year fixed effects. Based on the trend in figure 2.6, we divide the thirty years into 3 intervals: 1910-1915; 1915-1933; 1933-1940. In addition, we control for the total enrollment in science as it is the denominator of our outcome variable. As in the cross-section specifications, we include sizes of the student body as a control.

Columns 1 & 2 of table 2.8 show the results of the panel regression (2.2) & (2.3), respectively. Home Ec increases the fraction of women in science both on the extensive and intensive margins. The magnitude of the impact is substantial. After a Home Ec program opens, the fraction of women in science is expected to increase by 13.57 percentage points. A 4 percentage points increase in the Home Ec program's size is associated with a 1 percentage point increase in the fraction of women in science.

Columns 3 & 4 show estimates on a different outcome: the share of women in science. Because the reporting in the Arts & Sciences major increased over time, the number of students in science suffers under-counting more severely when the Home Ec dummy goes from 0 to 1. Thus, the estimates will understate the impact of having Home Ec programs. Even though the magnitude is not interpretable, the estimated coefficient of interests are positive and significant, indicating a positive spilled-over effect from Home Ec to science.

The above analysis excluded Cornell University. Although Home Ec was primarily a women's field, the gender reversed at Cornell University after 1930, thanks to the inauguration of the Hotel Management Program under the College of Home Economics. Even though hotel management was seen as a male occupation (Elias, 2008), the freshmen year curriculum required the typical science courses in Home Ec: hygiene, inorganic chemistry, and chemistry of food. Advised elective courses include general biology and physics (See figure 2.9). Could there be a parallel spilled-over to science for men? We regress the share of men in Home Ec on the fraction of men in science for all Cornell observations. Interestingly, the share of men in HE is positively correlated with the fraction of men in science. (See table 2.9)

The nature of compiling a dataset from numerous decentralized sources, where data content and format vary enormously even within the same source, means noisy measurement. We introduced several rounds of noises in preparing the yearbook data. We lost observations when students left the major field as blank or left their names in initials. We lost precision when the gender is inferred based on first names, and when the major field is too broad. Nevertheless, this data source has provided us with valuable evidence concerning women's dynamic enrollment in science in the early twentieth century. In future work, we plan to extend the data collection to private colleges and the period after 1940.

2.6 Mechanism

Why did Home Ec, a discipline closely tied to domesticity, increase women's enrollment in science? In this section, we provide some evidence on the potential mechanisms which underlie this finding.

Fundamentally, the spill-over effect was caused by overlapping in content, especially in the distribution requirements. In many ways, Home Ec studies developed as a feminine parallel to agricultural studies, so the requirements for domestic sciences around 1910 overlapped with those of farm sciences (chemistry and biological sciences).[12] For example, the 1919 Cornell Catalog shows that women who wished to specialize in Home Ec must complete the same core courses as men in agriculture, covering biology, chemistry, physics, physiology, and bacteriology. At Iowa State University, the first year courses for Home Ec were also identical with courses in agriculture (Eppright and Ferguson, 1971). Similarly, in the 1911 course catalog at Utah State University, students in domestic science must complete courses in general chemistry and plant physiology in freshman year. They must also take classes in bacteriology, advanced physiology and chemistry in sophomore year.

Through heavy exposure to sciences, Home Ec could lead women to a science degree in multiple ways. It reduced the amount of additional work to complete a double major in science. It could also inspire women to switch their majors entirely into sciences.[13] Moreover, it prepared women with all that was necessary to go to medical school or graduate school in science. There

---

[12] Although life sciences dominated Home Ec science requirements, there was an exception. When household equipment inaugurated as a Home Ec field in parallel to agricultural engineering, courses in physics and electric circuits were required (Bix, 2002).

[13] As long as the switching rate from Home Ec to science is less than 100%, switching will result in a positive correlation.

were plenty of anecdotes to support each channel. For example, Cassandra Wanzo went to Northwestern University in 1969 and majored in nutrition, a degree offered through the Department of Home Economics. She met all the requirements for a pre-med track and went on to medical school at the University of Wisconsin and became a psychiatrist in Atlanta (Blackwell, 2017). In another case, Reatha Clark King initially chose Home Ec major when she attended Clark College in Atlanta, considering a career in teaching Home Ec in high school. She fell in love with chemistry and switched her major to chemistry. She continued her education at the University of Chicago and completed her Ph.D. in thermochemistry (Spangenburg and Moser, 2003).

Data digitized from Student yearbooks helped us identify a pattern in the types of majors that were likely to pair with Home Ec. In all senses, double majors were rare. We observe a total of 56,314 undergraduates' declaration of majors, and only 648 (1.15%) of the undergraduates enrolled in more than one major. The percentage (2.7%) is higher for students in Home Ec: out of 5,179 students in Home Ec, 140 had a second major. Among the double-majors, 29 (20%) paired with education, and 45 (32%) paired with sciences.[14] In comparison, 171 out of 5,633 (3%) students who majored in education had a second major. Only 6 students had a double-major in science. [15]

2.7 Conclusion

In this paper, we study a historical pattern on American women's participation in science. Specifically, we argue that Home Ec programs expanded women's enrollment in science. We first present cross-sectional evidence between the size of Home Ec program and women's general science enrollment in 1910. We focused on the sample of land-grant universities because of their comparable institutional setup and academic standards. To generate exogenous variation in Home Ec enrollment, we use the demand for agricultural education as a source of identification. The empirical evidence supports the idea that exposure to the Home Ec curriculum increased women's participation in science.

---

[14] 35 general science, 3 industrial science, 1 science, 5 chemistry, and 1 zoology.

[15] 1 botany, 2 chemistry, 1 medicine, 1 physiology, 1 pharmacy.

The snapshot in 1910 captured a point in time when hygiene and sanitation was the central theme in Home Ec. This theme developed in the context of the prevailing germ theory and incorporated science courses such as bacteriology as its theoretical foundation. There was no evidence to suggest that women in Home Ec took a watered-down version of the science courses. Since Home Ec departments borrowed physical science courses from natural science departments or agricultural schools, Home Ec students were taught and evaluated in the same standard as their male classmates. Moreover, there was an incentive in the founding days to establish a rigorous scientific standard, as Home Ec strove for academic legitimacy.

We complement the cross-section analysis with a short run panel study on the opening of new Home Ec programs. Compared to when Home Ec was not available, the presence of Home Ec led to a higher proportion of women choosing a major in science and a substantial reduction in the science gender gap. Given that the Home Ec movement quickly spread to other private colleges and state universities, [16] extrapolating the panel estimates would imply a reasonably broad impact on women's entry into science.

Besides serving as a back door to science, Home Ec also invented important subfields in science. The first subfield is nutrition (or food science), the study of chemical substances relating to food. Students who graduate from food science can find employment as dietetics or food technologist. The second subfield is developmental science, the study of child behaviors and development. As Home Ec gradually phased out, developmental science became an integral branch in psychology.

Today, the Home Ec program is arguably irrelevant in the academic domain. In rare cases where this program continued, they are rephrased as either "Human Ecology" or "Family and Consumer Science." The Home Ec legacy is easy to overlook. Our paper examines the unique circumstances that gave rise to Home Ec, and highlights its relation to physical and biological sciences. This short-lived college program may have a long term consequence in women's fair representation in science today.

---

[16] The number of Home Ec programs increased from merely 68 in 1910 to 340 by 1970.

Table 2.1: Spill over to science : land grant universities

Women in general science

	(1)	(2)	(3)	(4)
Women in home ec	0.0993***	0.0808***	0.0986***	0.0996***
	(0.0205)	(0.0196)	(0.0216)	(0.0218)
Women in classics	-0.0095	-0.0164	-0.0072	-0.0099
	(0.0128)	(0.0121)	(0.0138)	(0.0132)
Number of majors	2.8443***	2.6917***	2.6224**	2.8599**
	(1.0317)	(0.9848)	(1.1604)	(1.0580)
Total students	-0.0024	-0.0053	-0.0037	-0.0022
	(0.0025)	(0.0035)	(0.0041)	(0.0027)
Scientific apparatus		4.1025		
		(3.1748)		
Teaching expenses on science		28.5038		
		(25.4913)		
Men in science		0.1042***		
		(0.0346)		
Total faculty			1.8801	
			(3.3518)	
Total books			-0.3537	
			(2.9093)	
Total endowment			-0.1536	
			(0.6598)	
Funds from public sources				-0.2365
				(1.0022)
Funds from private sources				-0.0520
				(0.3969)
Observations	48	48	48	48
R^2	.48	.61	.49	.48

Notes: The table reports OLS estimates. An observation is a land grant college in 1910. The dependent variable is the number of degrees in general science conferred to women. Women in home ec/classics is the number of degrees in general science conferred to women in home economics/classics. Number of majors is the total number of majors offered at the college. Scientific apparatus measures the value of scientific apparatus and teaching expenses on sciences measures the disbursement of funds on the teaching of natural sciences. Men in science equals to the number of degrees in general science conferred to men. Total faculty/books/endowment are measured in logarithm. Funds from public sources equals the amount of funds from state, federal government in logarithm. Funds from private sourses is measured in logarithm. University of California (Berkeley) is excluded because it was an anomaly in institutional setup: a merge between the private college of California and the public Agricultural, Mining, and Mechanical Arts College. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 2.2: The effect of funding sources on program sizes

Dependent Variable: # graduates in	Agriculture	HomeEc	Science	Engineering	Classical	Education
land grant	95.7574***	2.4963	-6.3288	195.7200***	-235.5455***	-13.2711
	(12.1969)	(9.1385)	(14.0454)	(31.6270)	(40.6233)	(24.9810)
land grant X public funds	0.0027***	0.0020***	-0.0003	-0.0010	-0.0004	-0.0014*
	(0.0004)	(0.0003)	(0.0005)	(0.0011)	(0.0014)	(0.0009)
landgrant X private funds	0.0012***	-0.0002***	-0.0003***	0.0025***	-0.0019***	-0.0005**
	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0004)	(0.0002)
size	Yes	Yes	Yes	Yes	Yes	Yes
state fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	573	573	573	573	573	573
R^2	.66	.28	.32	.59	.65	.26

Notes: The table reports OLS estimates. An observation is a college in 1910. The dependent variable is the number of degrees conferred in various major. land grant is a dummy if a college is a land grant university. Funds from public sources equals the amount of funds from state, federal government in logarithm. Size equals to the total number of degrees conferred at a college in 1910. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 2.3: Testing spill-over effect of other majors on science

#graduates in:	Women in science						Men in science
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Classics	-0.0282**						
	(0.0116)						
Music		-0.1097*					
		(0.0569)					
Education			-0.0389*				
			(0.0224)				
Fine Art				-0.0725			
				(0.0579)			
Commerce					0.1368		
					(0.2938)		
Household Economy						0.0896***	-0.0126
						(0.0203)	(0.0993)
Fundings	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Men in science	Yes	Yes	Yes	Yes	Yes	Yes	No
Women in science	No	No	No	No	No	No	Yes
Number of majors	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48	48	48	48	48	48	48
R^2	.39	.36	.35	.53	.31	.33	.28

Table 2.4: Spill over to science : First Stage & Reduced Form

	Women in home ec	Women in home ec	Women in general science	Women in general science
Men in Agriculture	0.1687*	0.1906**	0.0336**	0.0294*
	(0.0943)	(0.0934)	(0.0155)	(0.0146)
University Size Controls	Yes	Yes	Yes	Yes
Science related Controls	No	Yes	No	Yes
Observations	48	48	48	48
R^2	.24	.36	.23	.4

Notes: The table reports OLS estimates. An observation is a land grant college in 1910. The dependent variables in columns 1&2 are the number of degrees in home economics conferred to women. The dependent variables in columns 3&4 are the number of degrees in general science conferred to women. University size controls include the total number of majors offered and the total number of degrees conferred at the college. Science related measures include the value of scientific apparatus, teaching expenses on natural sciences, and the number of degrees in general science conferred to men. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 2.5: Placebo Test of Instrumental Variable on Various Outcomes

	Men in science	Teaching expenses on science	Scientific apparatus
Men in Agriculture	0.0322	1.0531	252.3759
	(0.0554)	(5.2379)	(174.4870)
University Size Controls	Yes	Yes	Yes
Funding Controls	Yes	Yes	Yes
Observations	48	48	48
R^2	.17	.11	.8

Notes: The table reports OLS estimates. An observation is a land grant college in 1910. The dependent variable in column1 is the number of degrees in general science conferred to men. The dependent variable in column2 is the teaching expenses on natural sciences. The dependent variable in column3 is the value of scientific apparatus. University size controls include the total number of majors offered and the total number of degrees conferred at the college. Funding controls include the amount of money received from public and private sources. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 2.6: Spill over to science : Second Stage

Women in general science	(2SLS)	(2SLS)	(OLS)	(OLS)
Women in home ec	0.1993**	0.1544**	0.1035***	0.0957***
	(0.0897)	(0.0652)	(0.0197)	(0.0194)
University Size Controls	Yes	Yes	Yes	Yes
Science related Controls	No	Yes	No	Yes
Observations	48	48	48	48
R^2	.19	.5	.48	.59

Notes: The table reports 2SLS and OLS estimates. An observation is a land grant college in 1910. The dependent variable is the number of degrees in general science conferred to women. Women in home ec is the number of degrees conferred to women in home economics. University size controls include the total number of majors offered and the total number of degrees conferred at the college. Science related measures include the value of scientific apparatus, teaching expenses on natural sciences, and the number of degrees in general science conferred to men. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 2.7: Yearbook Data Summary Statistics

College	Students Mean	Women Mean	Students in HomeEc Mean	Women in Science Mean	Yearbooks Num.	Yearbooks First	Yearbooks Last
Auburn University	343.88	66.13	25.25	0.63	8	1916	1940
Clemson University	238	23.6	0	1.8	5	1915	1940
Cornell University	732.53	260.79	20.5	11.74	38	1897	1936
Georgia School of Technology	274	44.18	0	0.53	17	1917	1940
Iowa State University	372.1	169.52	91.03	9.62	29	1905	1940
Louisiana State University	675.29	234	12.86	2.86	7	1927	1940
Missouri University of Science and Technology	74.42	27.08	0	0	12	1911	1940
North Carolina Agriculture and Technology	100	32	4	5	1	1939	1939
North Dakota State University	239	95.53	25.29	1.94	17	1908	1940
Texas Tech	378.5	163.5	50	2.5	2	1937	1940
University of Arizona	213.67	95.56	0	0	9	1913	1940
University of Colorado	218.93	100.78	0	0	27	1893	1939
University of Maine	208.76	83.88	8.12	5.36	25	1904	1940
University of Missouri	457.27	220.77	0.5	0.37	30	1905	1940
University of Nevada	74	28.57	3.71	2.86	7	1901	1940
University of New Hampshire	292.85	97.77	2.46	0	13	1909	1940
University of North Dakota	203.4	69.2	0	0	5	1906	1940
University of Washington	553	171	70	8	1	1940	1940
Utah State University	181.6	52.8	22.6	0.2	5	1911	1939
Virginia Tech	133.33	23.39	0	0.22	18	1898	1939
Washington State University	346.64	112.36	43.27	4	11	1903	1939

Notes: This is a list of colleges for which yearbooks are transcribed. For each college, we list the average number of students, average number of women, average number of students in Home Ec, average number of women in science with matched first names. Also listed is the total number of yearbooks transcribed, the earliest and the most recent transcribed yearbook

Table 2.8: Panel Estimates from Yearbooks Sample

	Frac. Women in science		Share Women in science	
	(1)	(2)	(3)	(4)
HE Exists	0.1357**		0.0242**	
	(0.0670)		(0.0105)	
Share of Women in HE		0.2584*		0.0523*
		(0.1364)		(0.0302)
Num. Women Students	No	No	Yes	Yes
Num. Science Students	Yes	Yes	No	No
Num. Total Students	Yes	Yes	No	No
Time intervals	Yes	Yes	Yes	Yes
Observations	122	122	217	217
R^2	.75	.74	.43	.43

Notes: The table reports OLS estimates. An observation is a college-year. The dependent variables in columns 1&2 are the estimated fraction of women science (women in science/total students in science). The dependent variables in columns 3&4 are the share of women in science (women in science/total women). Time intervals are dummy variables for years before 1915 and for years after 1933. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

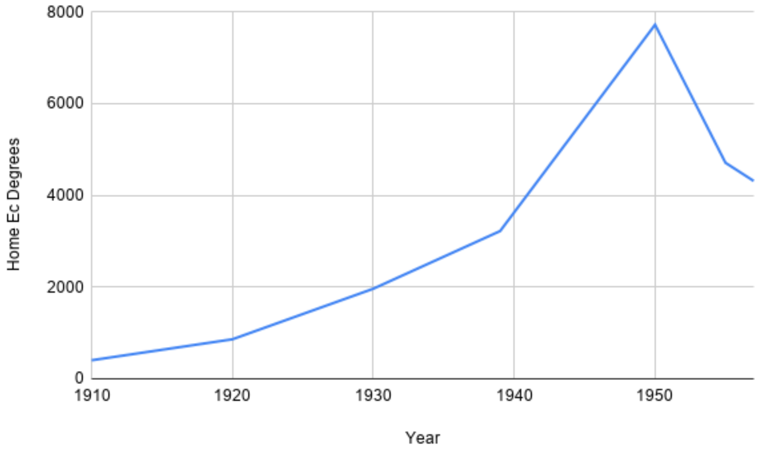
Table 2.9: Cornell University: Men in Home Ec and Science

	Frac. men in science	
	(1)	(2)
HE Exist	0.0821	
	(0.0481)	
Share of Men in HE		1.2012*
		(0.6190)
Num. Science Students	Yes	Yes
Num. Total Students	Yes	Yes
Time Intervals	Yes	Yes
Observations	23	23
R^2	.86	.87

Notes: The table reports OLS estimates. An observation is a year at Cornell University. The dependent variable is the estimated fraction of men science (men in science/total students in science). Time intervals are dummy variables for years before 1915 and for years after 1933. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Figure 2.1: Distribution of majors by institution type

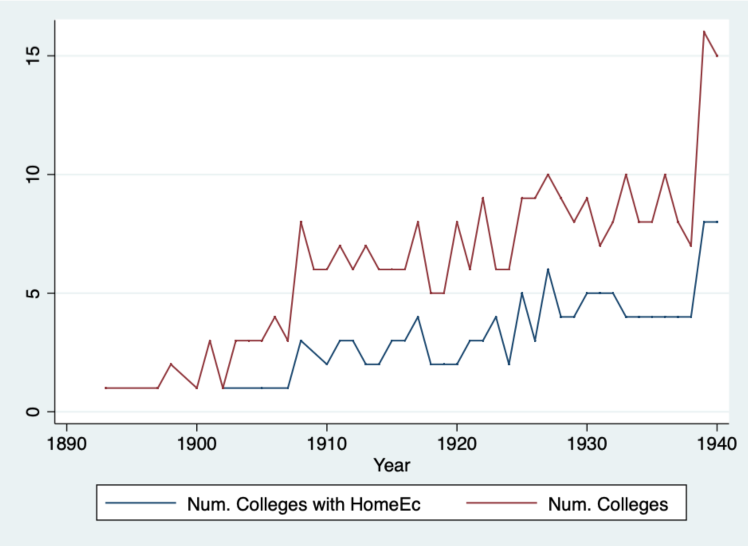
Figure 2.2: Number of Degrees in Home Ec conferred



Notes: Number of degrees in Home Ec conferred. Data source: Annual Report of the Commissioner of Education 1910; Biennial Survey of Education 1920, 1930, 1939, 1950, 1955, 1957.

Figure 2.3: Home Ec Trends from the yearbooks sample

(a) The number of colleges offering Home Ec



(b) Declared majors in Home Ec

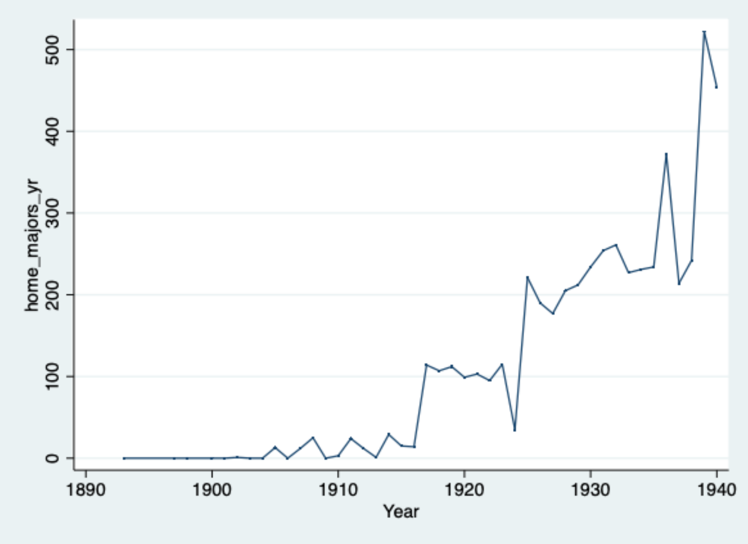
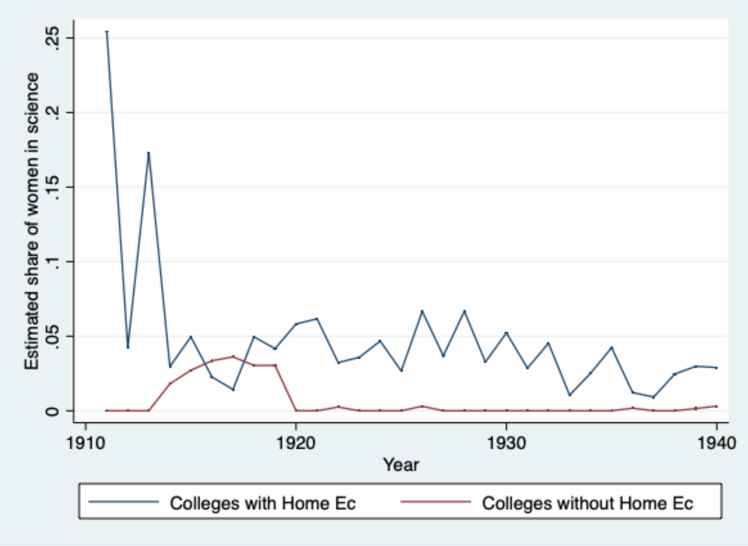
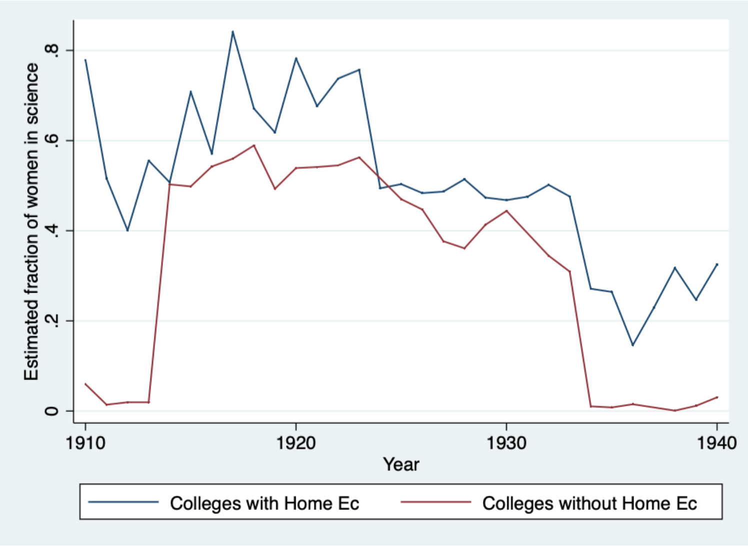


Figure 2.4: Share of women in science



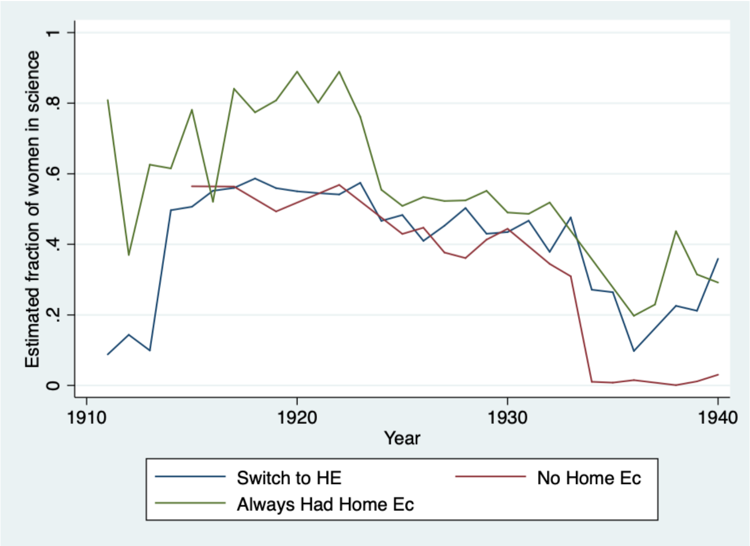
Notes: Unconditional mean share of women in science majors (women in science majors/total women) for colleges with Home Ec and colleges without Home Ec in each year. A student is counted as women if the first name has a probability of male less than 50% Data source:the student yearbooks sample.

Figure 2.5: Gender ratio in science



Notes: Unconditional mean fraction of women in science majors (men in science majors/ total science majors) for colleges with Home Ec and colleges without Home Ec in each year. Data source:the student yearbooks sample.

Figure 2.6: Gender ratio in science by treatment group



Notes: Unconditional mean fraction of women in science majors (men in science majors/ total science majors) for colleges that always had Home Ec, colleges that never had Home Ec, and colleges that switched to offering Home Ec. Data source:the student yearbooks sample.

Figure 2.7: Value of farm equipment and agriculture program size, 1910

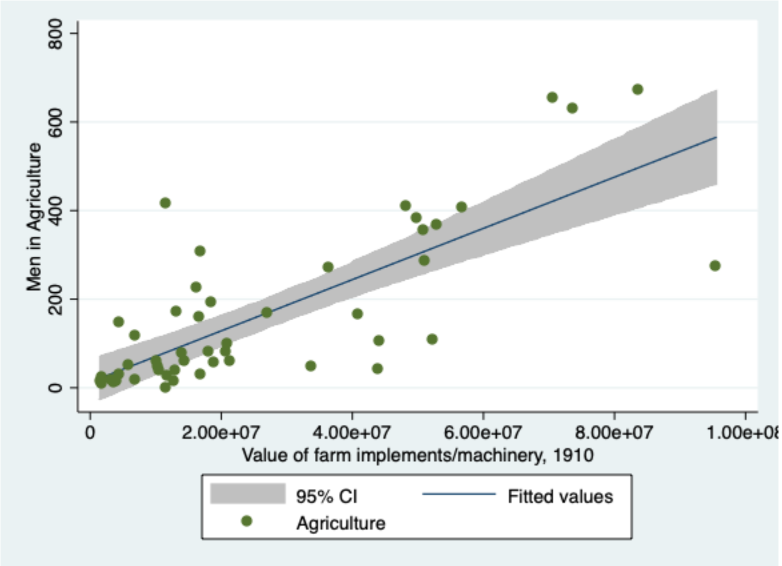


Figure 2.8: Home Ec and agriculture program sizes, 1910

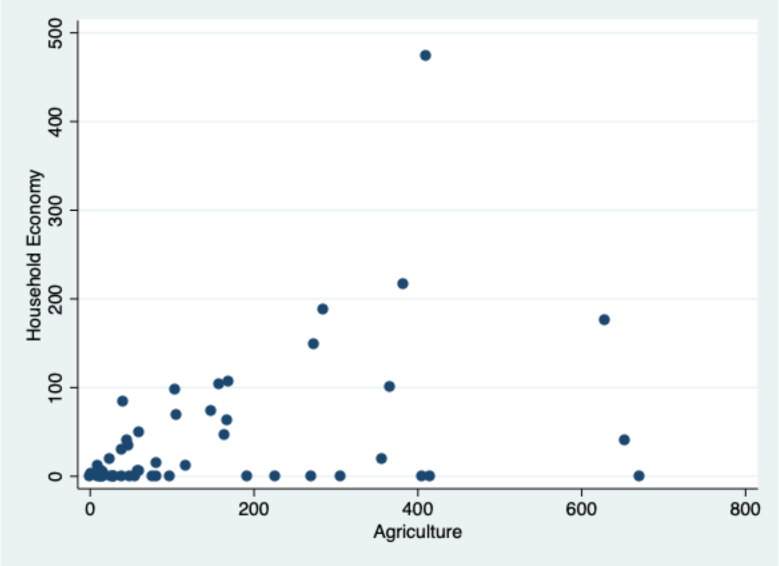
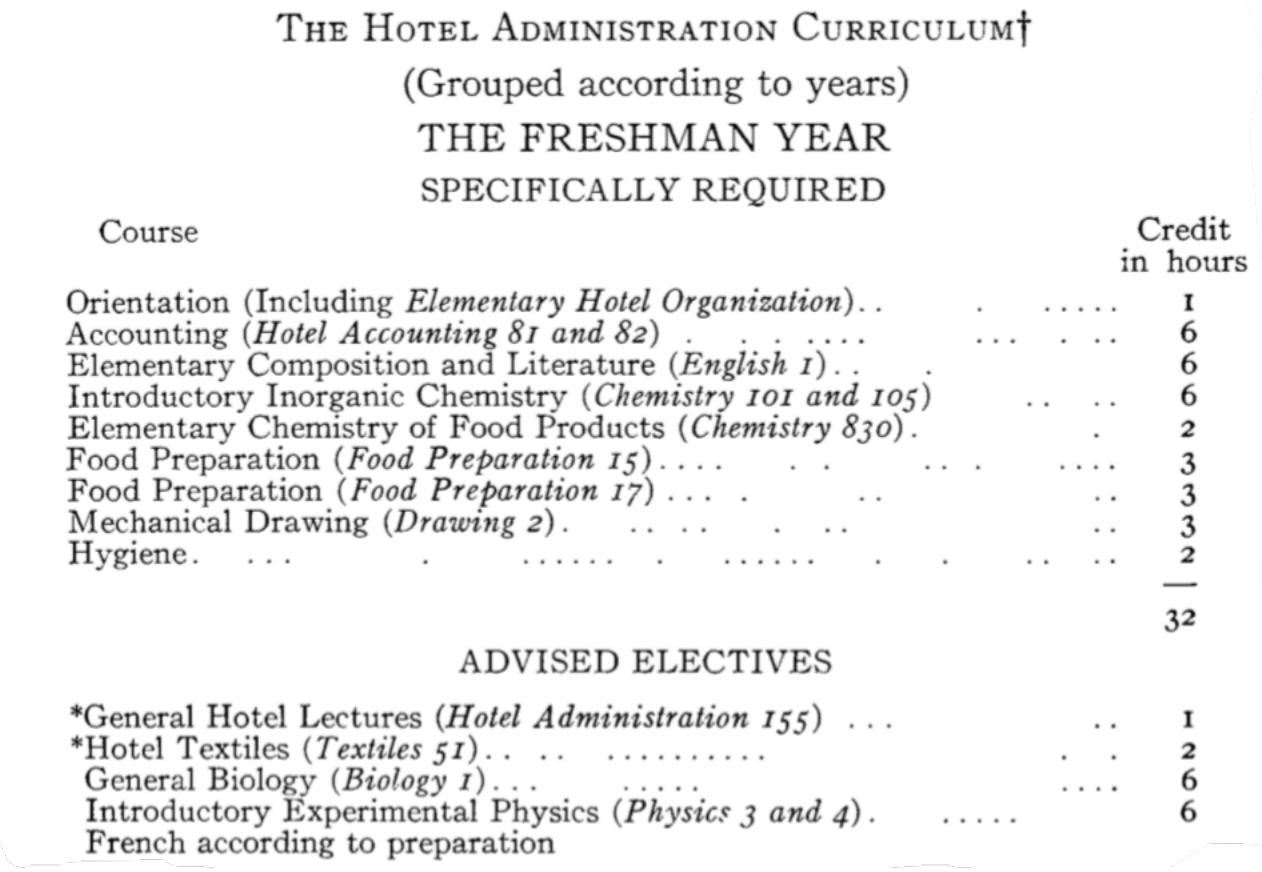


Figure 2.9: Cornell University Hotel Management first year courses



THE HOTEL ADMINISTRATION CURRICULUM†

(Grouped according to years)

THE FRESHMAN YEAR

SPECIFICALLY REQUIRED

Course	Credit in hours
Orientation (Including *Elementary Hotel Organization*)	1
Accounting (*Hotel Accounting 81 and 82*)	6
Elementary Composition and Literature (*English 1*)	6
Introductory Inorganic Chemistry (*Chemistry 101 and 105*)	6
Elementary Chemistry of Food Products (*Chemistry 830*)	2
Food Preparation (*Food Preparation 15*)	3
Food Preparation (*Food Preparation 17*)	3
Mechanical Drawing (*Drawing 2*)	3
Hygiene	2
	32

ADVISED ELECTIVES

*General Hotel Lectures (*Hotel Administration 155*)	1
*Hotel Textiles (*Textiles 51*)	2
General Biology (*Biology 1*)	6
Introductory Experimental Physics (*Physics 3 and 4*)	6
French according to preparation	

CHAPTER 3

ENGINEERING THE GENDER GAP: FALL OF WOMEN'S SHARE IN COMPUTER SCIENCE

3.1 Introduction

Since World War II, women have made considerable progress in both labor force participation and college enrollment. Women now receive 57 percent of all bachelor's degrees in the United States, an increase from 38 percent in 1965.[1] Despite the increasing level of participation, there is a pattern to women's academic preferences. It is fairly well known that women are less likely to major in STEM fields: they make up 35.5% of all recipients of STEM bachelor's degrees. Within STEM, women are far less likely to choose engineering (19.5%) over math and science (53%). If we break down science, women are the majority in the life sciences (59.9%) in contrast to the physical sciences (38.5%).[2] At a time when women receive more higher education than men and are free to pursue any area of study, what contributes to this pattern of gendered human capital accumulation?

Scholars have tried to explain the gender gap in STEM based on gender differences in math ability. But numerous studies show that the gender gap in math achievement and aptitude is small and diminishing (Goldin and Kuziemko, 2006; Reuben et al., 2014). Additionally, math and science performances in primary and secondary levels predict poorly whether a student majors in STEM in college (Lubinski and Benbow, 1992; Stoet and Geary, 2018; Turner and Bowen, 1999; Xie and Shauman, 2003). Math ability cannot explain either why women prefer math over engineering, or life sciences over physical sciences.

---

[1] See figure C.7 for the share of women earning bachelor's degrees from 1965-2016. The numbers are computed from the IPEDS survey (Integrated Postsecondary Education Data System, 1986–2016) and the HEGIS survey (Higher Education General Information Survey, 1965–1984).

[2] The numbers in parentheses are the percentages of women who graduated in the corresponding fields.

In this paper, I argue that traditional gender roles in the U.S. influence womens academic preferences today. Historically, women were given a gendered education, focused on subjects that served their gender roles. Some STEM fields were a part of womens education because, for example, math allowed them to help their husbands keep accounts, and science was meant to prevent superstition (Rush, 1787). In the late nineteenth century, psychology, chemistry, biology, and health were incorporated into womens education through collegiate home economics, a field developed around women's domestic duties. These scientific fields have the highest percentages of women majoring in them today. On the other hand, some disciplines never found immediate domestic applications and remained male dominated. I use computer science (CS) as a case study to show that academic norms rooted in traditional gender roles still influence womens choices today: when this major moves from colleges of liberal arts & sciences to the traditionally masculine domains of schools of engineering, the percentages of women in CS falls.

One might think that traditional gender roles have become less relevant in the 21st century, but much evidence indicates their prevalence. The gender division in a traditional American family, where the husband is responsible for the public sphere and the wife is responsible for the private (e.g., keeping a clean home, managing household expenses, raising children), is still mainstream. Men continue to prefer female partners who are less intelligent and professionally ambitious than they are (Fisman et al., 2006). In a field experiment, single female MBA students sacrificed career-enhancing opportunities, such as the expectation to work long hours and frequent traveling, to avoid signaling unmarriageable traits in the presence of male classmates (Bursztyn et al., 2017). A marriage where the woman earns the majority of household income or takes a leadership position is still unlikely, and when that happens, marital satisfaction is lower and the divorce rate is higher (Bertrand et al., 2015; Folke and Rickne, 3002forthcoming). It is reasonable to believe that a forward-looking woman (or her parents) who is interested in a mainstream marriage will face a tradeoff between pursuing college majors of her passion and ambition and investing in college majors in accordance with traditional gender roles.

While traditional gender roles persist, the substance of homemaking has evolved to be more heterogeneous over time, thanks to historical development in cultural norms and scientific research that continued to shape the private sphere. Three significant cultural shifts and scientific discoveries that transformed homemaking took place in the late nineteenth century and early twentieth century (Mokyr, 2000). The first was a shift in the belief that children deserved protection, nurture, and education. The second was a breakthrough in bacteriology. Scientists came to understand that germs caused disease and infection, yet hygiene and sanitation prevented them. The third was a discovery that linked nutrition, such as vitamins and minerals, to health. The realization that disease was controllable and preventable through domestic cleanliness and adequate nutrition shifted the responsibility for the well-being of household members from "fate" to the homemakers. Because homemakers were primarily women, they were disproportionately affected. Since the 1870s, they were found to be working longer hours in their homes, despite the growing availability of labor-saving household appliances (Cowan, 1985).

The creation of collegiate home economics reflected changes in homemaking by formalizing modern domestic topics into its subfields of *Child Care & Human Development*, *Hygiene*, and *Food & Nutrition*. Using titles of books and journals in home economics literature published between 1800 to 1950,[3] I document that writings in these three topics indeed surged around the late nineteenth century and early twentieth century. In terms of content, home economics was more closely related to STEM than we realize today. By tabulating the frequency of science subjects and their related terms that appear in the titles, I map the correspondence between science subjects and each domestic topic: psychology was incorporated into *Child Care & Human Development*; chemistry into *Food & Nutrition*; and biology and public health into *Hygiene*.

Even though designed with domestic purposes, home economics had broader implications beyond the family home. By applying modern sciences into its curriculum and professionalizing what women used to do at home, it opened up many academic and professional avenues for women that were still agreeable to the traditional female roles. Today, science subjects that are closely related

---

[3] Data source: Home Economics Archive: Research, Tradition and History (HEARTH) (Albert R. Mann Library, 2005)

to home economics have the highest shares of women. The persistence effect of home economics may arrive from two channels: (i) after home economics was phased out in the 1970s & 1980s, women who were interested in domesticity substituted home economics with its related science subjects; (ii) women who were genuinely interested in these science subjects no longer faced the trade off with traditional gender roles.

I test the substitution mechanism by comparing shares of women in sciences after a university re-opened a home economics department after the 1980s. I look at opening effects instead of closing's because closing a home economics department is endogenous to a decreased demand for this major. An event study on currently enrolled cohorts shows that when universities re-started offering home economics (now often disguised as human ecology or family and consumer sciences), share of undergraduate women in biology, chemistry, and psychology are crowded out. There is no change in the share of women going into engineering, physics, or geoscience, indicating that the substitution of major is related to the curriculum of home economics.

Without directly addressing the second persistence channel in home economics, I test the trade-off between academic interests and traditional gender roles in a different context: computer science (CS). I take advantage of the fact that engineering, with its applications to industry and military, is far more distant to the traditional female role than math or science ever was. And CS is a major that can be associated with either engineering, or math and science.[4] In addition to the cross-sectional variation, CS strengthened its association with engineering over time: among universities with engineering schools, the percentage of CS programs organized within which has increased from roughly 45% in 1980 to 75% in 2010.[5] I use within university variation in CS departmental affiliations to test the hypothesis that the proportion of women studying CS decreases as CS moves from colleges of liberal arts & sciences to schools of engineering.

---

[4] CS has a strong lineage in mathematics: it traces to mathematicians like Alan Turing and John von Neumann; many CS programs were initiated under math departments. CS is also offered through many engineering schools, often times under electrical engineering departments.

[5] CS used to have a larger share of women (37% in 1984) and now it is a more male-dominated field (19% in 2016). The fraction of women in CS dropped to a level comparable to that in engineering (20.9% in 2016).

The data on CS departmental affiliation–the school within which CS program is placed–is based on information from the Course Catalog Survey (Brint, 2013), a database on university hierarchy covering 286 institutions at 5 year intervals, including 243 colleges with CS programs. The database records department name, name of the residing school, and their corresponding fields and disciplines. Combining information on school names and their disciplines, I specify whether a university affiliates its CS program with engineering and whether or not this affiliation is exclusive.

I then link these institutions to the Degree Completion Surveys (Higher Education General Information Survey, 1965–1984; Integrated Postsecondary Education Data System, 1986–2016), which report the number of degrees received by field, gender, and degree level. Ultimately, I assemble a college panel dataset from 1980 to 2010 with both CS department affiliation information and degree completion by gender, at 5 year intervals.

The main empirical strategy is a fixed effect model. I compare CS programs across affiliations within the same university, thereby holding constant many time-invariant factors that determine an engineering affiliation, including public vs. private status, liberal arts vs. comprehensive curricula, selectivity, location, and the heritage in professional education. Controlling for year fixed effects allow me to employ randomness in the timing of re-affiliation decisions, thereby accounting for development in the underlying IT market that can impact both re-affiliation decisions and gender composition in CS. The primary analysis reveals a significant and negative relationship between engineering affiliation and the proportion of women in CS. When a university moves its CS program to the school of engineering exclusively, the fraction of women in CS decreases by 15 percent. Given that the fraction of women in CS has dropped 50 percent from its peak, my finding is quantitatively meaningful.

To support the interpretation of the engineering affiliation effect as a constraint from traditional gender roles, I conduct several tests to rule out potential alternative factors that might generate the result, such as admission cost, masculine culture, and changes in degree requirements. After all, even different schools within the same university can vary in culture, resources, ranking, and requirements. I acknowledge that moving the CS program to the engineering school incurred more than a change in the degree of male association.

I address the concern over switching costs - costs associated with applying to the engineering school and satisfying the different general distribution requirements for engineering - by comparing CS programs that become a part of electrical engineering and computer science (EECS) department to CS programs that become an independent major in the engineering school. Because EECS department locates in the engineering school, it has the same admission and the same general distribution requirements as other engineering majors. If switching costs drive the result, we should not expect differentiated responses from the additional affiliation with EE. However, I find a further drop in the proportion of women in CS programs that locate to EECS. Given that EE is more male-dominated than other engineering fields, such as textile, biomedical or chemical engineering, this finding not only disproves switching costs as the primary mechanism, but also gives support to the proposed interpretation.

Next, I address the fact that placing CS in the engineering school creates not only a symbolic distance to the traditional female roles but also a tangible masculine culture known to marginalize women and harm their productivity (Banchefsky and Park, 2018; Cheryan et al., 2017; Walton et al., 2015). The former channel offsets women's interest to enroll in a CS program, whereas the latter hurts the survival rate of already enrolled women. To isolate the two channels, I explore variation across universities in how male-dominated their engineering programs are, relative to their CS programs before re-affiliation. While masculine culture is a significant factor–the more male-dominated the engineering school into which a CS program moves is, the greater the drop in the fraction of women in CS–controlling for this gender gap does not crowd out the estimated coefficient on engineering affiliation.

There is also the concern that relocating the CS program to engineering schools involves significantly more coursework in engineering. To alleviate this concern, I collected CS degree requirements from a sample of university course catalogs. I transcribed the maximum number of math, CS, and engineering courses that counted towards a CS degree in the minimum standard.

On average, a CS degree only required one additional engineering course from 1970 to 2000. This contrasts with a proximately 30 percent increase in the share of engineering affiliations. Except for Virginia Tech and Southern Illinois University, both of which added two engineering courses, none of the universities in my sample that relocated CS to their engineering schools added additional engineering courses to their degree requirements.[6] However, I acknowledge that changes in course load do not reflect changes in difficulty, nor does it rule out the expectation that the same courses will be more difficult in the engineering school.

Finally, the negative relationship between the fraction of women and engineering affiliation only exists at the undergraduate level. There is no evidence that re-affiliation affected CS students from masters programs. Because foreign male and female students constitute more than half of the male and female CS graduates from masters programs (See figure C.5) , their dominance suggests that the gendered academic preferences towards math and engineering are specific to gender norms formed in the U.S. context.

Among domestic undergraduates, the magnitude of the gender gap in CS is the largest among white women. In 2016, they constituted 45% of female CS graduates but 59% of all bachelors degrees awarded to women. Black women, in contrast, are well-represented. They constituted 11.4% of female bachelors degree holders in CS and 10.5% of female degree recipients across all fields. Asian women are over-represented among CS bachelors degree recipients as they constituted 15.6% of female CS graduates, relative to 6.3% of female college graduates. The racial differences are not accounted for by differences in technical preferences: white women are perfectly well represented in math and science. Confirming my hypothesis, white women's under-representation in CS is consistent with their under-representation in engineering, and electrical engineering in particular.

This paper relates to the broader literature in gender economics. Gender differences in major continue to explain a large share of the existing wage gap among college educated men and

---

[6] The other universities that relocated CS to engineering in my sample are Clemson University, State University of New York at Buffalo, Pennsylvania State University, University of California-Irvine, University of Rochester, Ohio University, University of Nebraska Omaha, and West Virginia University.

women (Dan A. Black and Taylor, 2008). Other than gender differences in math achievement, scholars have been searching for alternative explanations for these differences. In a systematic attempt, Wiswall and Zafar (2015) accounted for gender differences in undergraduate majors by the systematic differences in male and female priorities. Their finding that male students care more about economic return is in keeping with the traditional gender division where men are seen as the breadwinners in marriage. Stoet and Geary (2018) confronts the paradox that in less gender-equal countries, there are larger shares of women major in STEM despite more pronounced gender gaps in math scores. This behavior is rationalizable: when pressure on quality of life is high and men cannot be relied on to be the sole breadwinners, traditional gender division becomes less appealing than the economic return associated with STEM. My paper brings in additional evidence to support the idea that the gendered pattern of human capital accumulation is rooted in traditional gender roles.

This paper is organized as follows. Section 2 provides historical background. Section 3 discusses the related literature. Section 4 describes the data, followed by results in Section 5. Section 6 concludes the paper.

3.2 Historical background

3.2.1 From big irons to minicomputers

The percentage of women in computer science started falling at approximately the same point in time when personal computers started replacing mainframe computers. This is not a coincidence. The advent of personal computer marked a hardware revolution with a direct consequence in the underlying labor market that disproportionately affected women.

Early mainframe infrastructure occupied a room, referred to as a data center, and shared space with many other subsidiary hardware components for input, output, and memory.[7] The mainframe computer consisted of a main storage device, a system console, central processing units (CPU), peripheral processors, and input/out devices (I/O). I/O included such equipment as card readers and punches, magnetic tape units, disk storage, drum storage, typewriter-keyboard devices, printers, teleprocessing devices, and sensor-based equipment (International Business Machines Corporation, 1974). Users connected to the mainframe using terminals and submitted their tasks for processing by the mainframe.

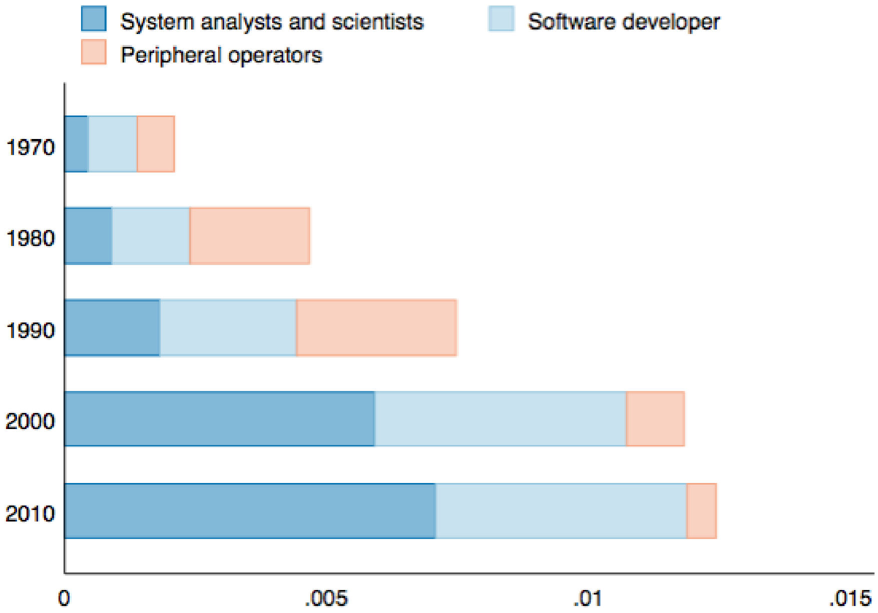
A large data center required technical support to allow the IT infrastructure to function. The computer operators (or peripheral operators because they worked on the peripheral parts of the mainframe) held the jobs associated with running the data center. Figure 3.1 shows there were more peripheral operators in the computer industry than programmers in 1980 and 1990.[8] Figure 3.2 shows that more than half of the peripheral operators were women. As mainframe computers were phased out, it eliminated the need for numerous computer operators to run mainframe systems. In addition, functions previously performed by computer operators were increasingly being done by other computer workers (U.S. Dept. of Labor, Bureau of Labor Statistics, 2000). Women's share in the computer industry thus plummeted. Figure C.3 shows the percentage of women in the computer industry by using the 1990 occupational category defined by the census.

---

[7] Picture C.1 shows an IBM mainframe computer.

[8] This graph is generated from IPUMS random population samples. According to Occupational projections and training data provided by U.S. Dept. of Labor, Bureau of Labor Statistics (1976), there were 233,000 computer operators jobs in 1980 and 247,000 programmers employed in 1978.

Figure 3.1: Growth in computer related occupations: 1970-2010



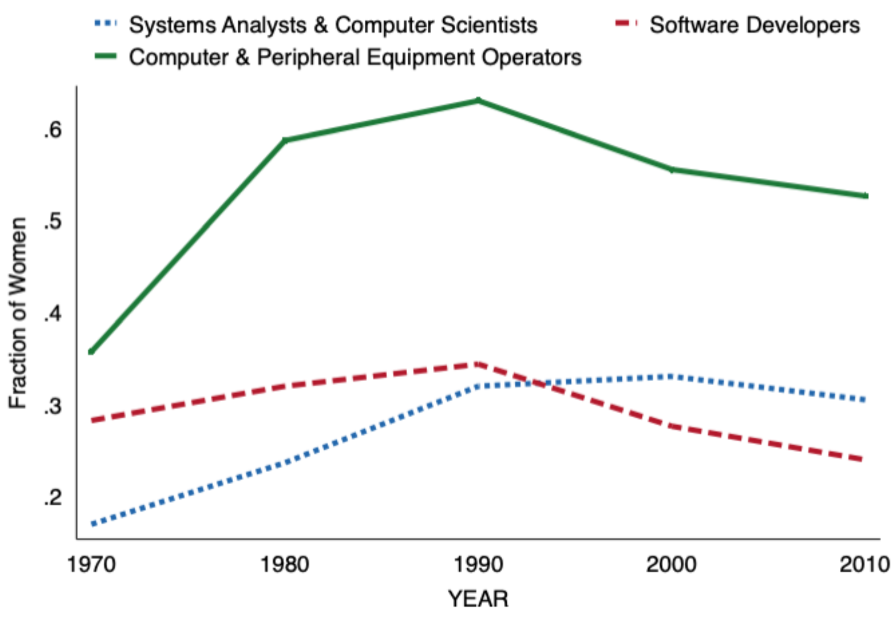
Notes: This figure shows the relative share of population that are employed as system analysts & computer scientists, software developers and peripheral operators. *Data sources*: IPUMS random population sample: 1970 1% state fm1, 1980 5% state, 1990 5% state, 2000 5% state, 2010 ACS (Steven Ruggles and Sobek, 2019)

However, computer operator jobs were not completely segregated from programmer jobs. Wages for advanced operators were comparable to those for lower level programmers (Prieser, 1985). Hiring was not all that different from hiring for programmers, as many employers tested computer operator applicants to determine their aptitude for computer work, particularly their ability to reason logically. During the year 1980, of all employed computer operators who left the occupation, one-fourth of them advanced to the position of computer programmer (U.S. Dept. of Labor, Bureau of Labor Statistics, 1976).

Schultz (1975) insightfully proposed that returns to human capital were highest in economic disequilibria, if one can adapt resource allocations to those shocks. So what then prevented women from transitioning to becoming programmers after the 1980s? What constrained women's ability to reallocate their human capital to respond to this technology shock?

The reason could be that programming had become increasingly associated with engineering in both industry and academia. In the workforce, the most popular job title for programmers and

Figure 3.2: Women's share by year and occupation



Notes: This figure shows the fraction of women that are employed as system analysts & computer scientists, software developers and peripheral operators. *Data sources*: IPUMS random population sample: 1970 1% state fm1, 1980 5% state, 1990 5% state, 2000 5% state, 2010 ACS (Steven Ruggles and Sobek, 2019)

developers is now "software engineer". In universities, CS majors are increasingly likely to be located in the engineering school. Since engineers and engineering programs are predominantly male, this gender norm may diminish the likelihood that women stay in the computer industry or enroll in a CS major.[9]

In subsection 2.2, I provide evidence suggesting that "software engineer" has taken place of "programmer" in popularity. In subsection 2.3, I look for historical explanations for why engineering has been accepted as more masculine than science. Finally, I argue that what happens in the universities parallels what is happening in industry, and analyze this university setting in sections 4 & 5.

---

[9] In the labor force, only 15.9% of engineers are women in 2019. https://www.bls.gov/cps/cpsaat11.htm

3.2.2 From programmer to software engineer

Prior to the mid-1970s, software practitioners were generally called computer programmers or software developers. It was widely agreed what those terms meant, while the exact meaning of software engineer was still being debated.

The term “software engineering” was first cited as the title of a 1968 NATO-sponsored conference on software engineering in Garmisch, Germany. It was also said to have been coined by Margaret Hamilton, a women software developer in the Apollo program (Rayl, 2008). Regardless of its origin, the common goal was to legitimize software development as similar to other engineering disciplines.

Academics in computer science disagreed vehemently on this term. In 1978, computer scientist E. W. Dijkstra argued in an essay that the coining of the term “software engineer” would lock in minimal standards.

> The existence of the mere term has been the base of a number of extremely shallow– and false– analogies, which just confuse the issue... Computers are such exceptional gadgets that there is good reason to assume that most analogies with other disciplines are too shallow to be of any positive value, are even so shallow that they are only confusing. I can only conclude that the wide-spread adoption of the term “software engineering” is to be regretted as it only hampers this recognition (Edsger W. Dijkstra, 1978).

Dijkstra represented the point of view held by computer scientists trained in mathematics. Engineering faculties, on the other hand, embraced the idea of software engineering. David Parnas, a computer scientist trained in electrical engineering, advocated the modeling of software engineering on programs in traditional engineering discipline to “increase both the quality and quantity of graduates who are well prepared, by their education, to develop trustworthy software products” (Parnas, 1998).

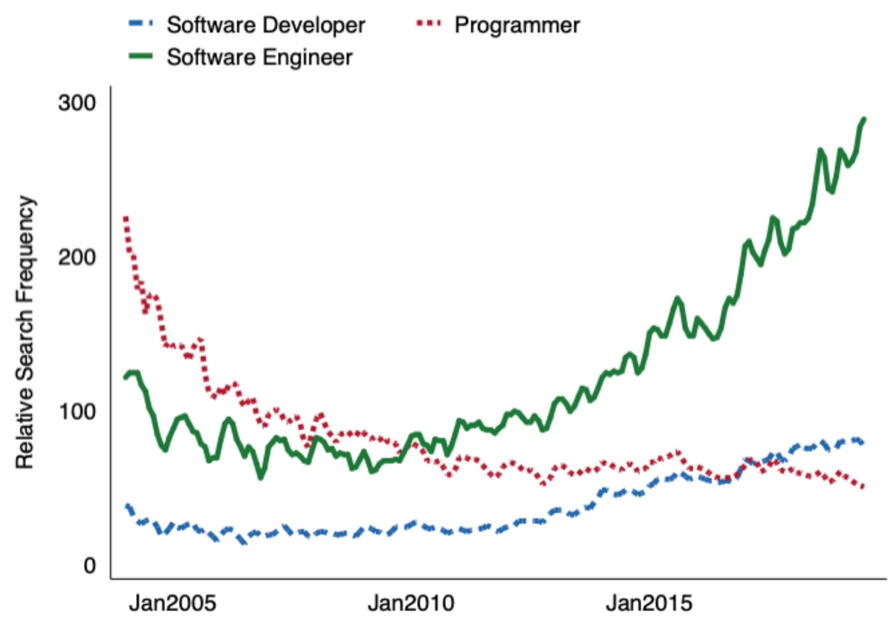
The rise in software engineering was grounded in changes in industry demand. In the mainframe era, programming provided customized solutions to government and large businesses for data processing: payroll, airline reservations, credit card transactions, etc. (R.Yost, 2017). This led to a focus on computerized assistance.[10] With personal computers, programmers could apply their skills to develop applications and create products, which resembles an engineering process.

It took more than 40 years for software engineering to mature into an academic discipline. In 2005, a guide book on the commonly accepted software engineering body of knowledge, known by the acronym SWEBOK, was finally developed. This publication officially set standards for software engineering, and following this, software practitioners were increasingly called software engineers instead of programmers or software developers. This trend is reflected in the relative search frequencies for these terms by Google users in the U.S. (figure 3.3). Among the searches relating to jobs, the popularity of "programmer" decreased sharply over time. "Software engineer" dominated both "programmer" and "software developer" in search frequency in approximately 2010 and it has become the job title we associate with programmers today.

---

[10] To computerize a payroll system, for instance, an applications developer had to interview everyone currently involved in the payroll process, understand and document their contributions to the process in explicit detail-not failing to account for exceptional cases and infrequent variations in normal procedures-and then translate these complex activities first into a form that other programmers could understand and eventually into the precise commands required by the computer. Since the payroll department did not operate in isolation, they had to work with other departments to coordinate activities, standardize the required inputs and outputs to the procedures, and negotiate points of conflict and contention. They also had to produce documentation, train users, arrange testing and verification procedures, and manage the logistics of implementation and roll-out (Ensmenger, 2010) (p.269).

Figure 3.3: Programmer, Software developer & Software Engineering



Notes: This figures plots the relative search frequencies in the job category on "software developer", "programmer", "software engineering" by Google users in the U.S. The frequencies are averaged over a three year moving window. *Data source*: https://trends.google.com

3.2.3 From math to engineering

Education in the United States has a gendered history. Women were initially educated to become dutiful wives and mothers, an ideal rendered some subjects to be more appropriate than others. The legacy has persisted through time and manifests itself in women's choices of college majors today. In this subsection, I narrow my discussion to STEM majors and provide reasons for why math and sciences are less male-dominated than engineering.

Contrary to the common misconception, math has always been a gender neutral subject. As late as the 1960s, math at the undergraduate level had the same percentage of women, approximately 35 percent, as were enrolled in universities overall. Engineering, on the other hand, had virtually no women.[11]

---

[11] See figure C.9.

Math and sciences were long thought to be more valuable to domesticity than engineering. When higher education was not widely accessible to women, math and sciences were introduced to girls from upper- and upper-middle-class families, providing they did not take their scholarship seriously. The reason, as educator Benjamin Rush spelled out in his 1787 speech to the Young Ladies' Academy of Philadelphia, was so that "they could maintain household budgets and help a husband with business...and to prevent superstition (Rush, 1787)." Engineering, meanwhile, was not remotely related to women's domestic duties. In the nineteenth century U.S., engineering was associated with hard, physical labor and projects such as building the Erie Canal, railroads and mines. The attitudes about women in engineering were reflected in the statement made by the Dean of the School of Engineering at Penn State University in 1955, that "...women are not suited to engineering, and that teaching them is wasted effort" (Bix, 2014).

State and land-grant universities made higher education accessible to many more women, particularly to those from rural areas. However, co-education still implied different academic objectives for men and women: for men, to be trained in the newest methods of farming or engineering; for women, to learn new ideas about how to perform housework. Trustees at Iowa State College, which admitted women from its opening in 1869, declared, "if young men are to be educated to fit them for successful, intelligent & practical farmers & mechanics, is it not as essential that young women should be educated in a manner that will qualify them to properly understand and discharge their duties as wives of farmers & mechanics?" To achieve this goal, they increased facilities for scientific instruction, with a "prominent place" for "the study of domestic economy".

Even though home economics was understood to be "furnishing wives", it introduced science to working class women. The first homemaking class at Iowa State was "Chemistry as Applied to Domestic Economy". Aside from one course in "ladies' course of study", the rest of the education required female students to take a years worth of inorganic and organic chemistry or qualitative analysis, botany and physics, psychology, comparative anatomy and physiology, geology, or meteorology.

Land grant universities led private universities in the movement towards home economics. Even though this was not a professional degree, generations of women were introduced to chemistry, biology, and psychology via courses on “Chemistry as Applied to Domestic Economy”,“Microbiology and canned food” and “The psychology of childhood”. I document word frequency of different science subjects from Home Economics bibliography (Albert R. Mann Library, 2005) as presented in chart 3.6. Psychology-related terms such as “psychology”, “psychiatric” and “mental” appeared most frequently, occurred under the category on *Child Care, Human Development, & Family Studies Bibliography*. Biology and chemistry-related terms were common-place under the category *Hygiene & sanitation* and *Food & Nutrition*. Figure C.18 presents the share of women in various science subjects and over time. Women dominate in psychology: an average psychology program in the U.S. had close to 80 percent women in 2016. Biology ranks second in popularity with 60 percent of women, slightly leading slightly ahead on chemistry which has an equal ratio of women to men. The percentage of women in mathematics programs has shown a rather steady trend over time, fluctuating around 40 percent.

These trends demonstrate that math and science have been a part of American womens education for more than 100 years. Today, women make up nearly half of these fields, if not more. In contrast, engineering has never been targeted towards women.

3.2.4 Gender and Race

To empirically test the relationship between home economics and women’s participation in sciences, I explore within university variation and compare the percentage of women in sciences after a university opened home economics major. Columns 1 & 2 of table 3.7 show that home economics crowds out the share of women in physical science (PHSC), but not in engineering.

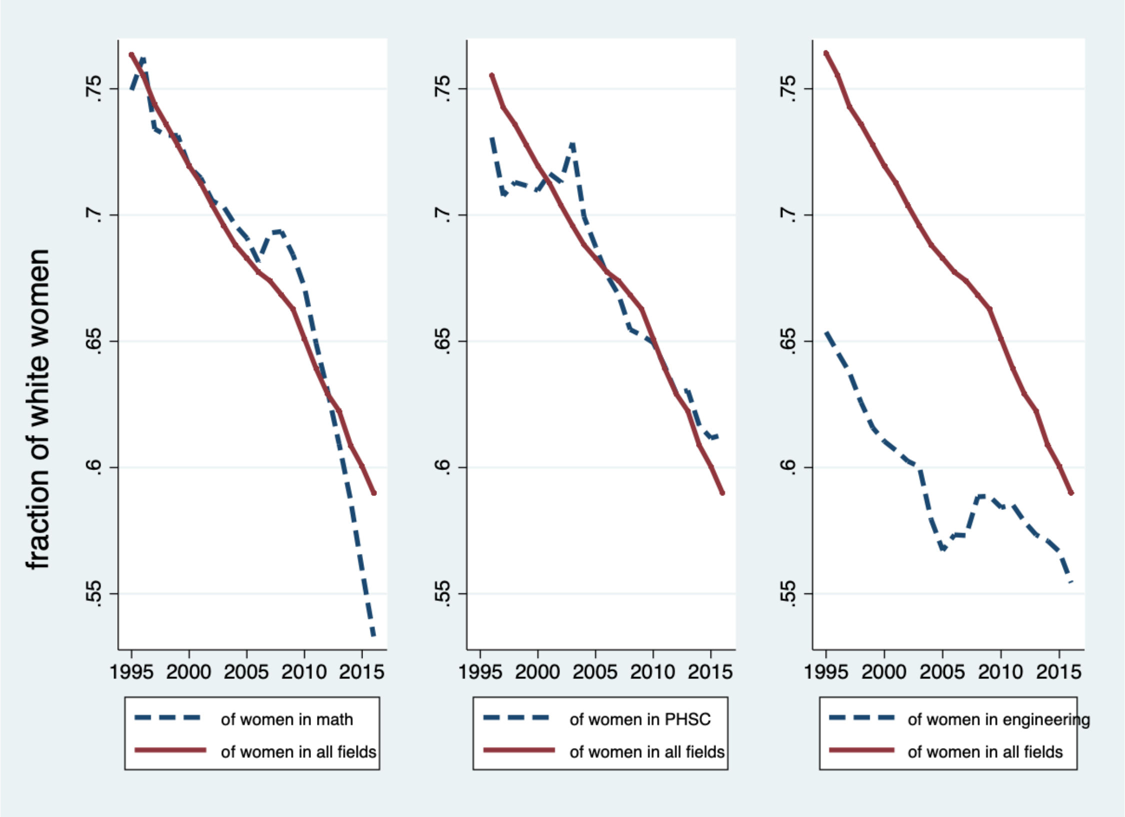
White women over-represent in home economics compare to other races. In 1995,[12] white women constituted 83% of female home economics graduates, relative to 75% of female college graduates across all fields. While women from other races are either under- or over-represented

---

[12] 1995 is the first year race statistics were reported from (Integrated Postsecondary Education Data System, 1986–2016).

in both math and engineering, white women demonstrate a pattern specific to the gender norm described in the previous subsection. They are well-represented in math and physical sciences, yet they are under-represented in engineering (figure 3.4).

Figure 3.4: White women in math, physical science and engineering



Notes: This figure depicts white women's relative representation in math, physical science and engineering. The solid lines plot the fraction of white women among female college graduates across all fields. The dashed lines plot the fraction of white women among female college graduates in math, physical science or engineering. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

Confirming my hypothesis, white women's under-representation in engineering mirrors their under-representation in CS. By 2016, the magnitude of the gender gap in CS is the largest among white women. They constituted 45% of female CS graduates but earned 59% of all bachelors degrees awarded to women. Black women, in contrast, are well-represented. They constituted 11.4% of female bachelors degree holders in CS and 10.5% of female degree recipients across all fields. Asian women are over-represented among CS bachelors degree recipients as they constituted 15.6% of female CS graduates, relative to 6.3% of female college graduates across all fields.

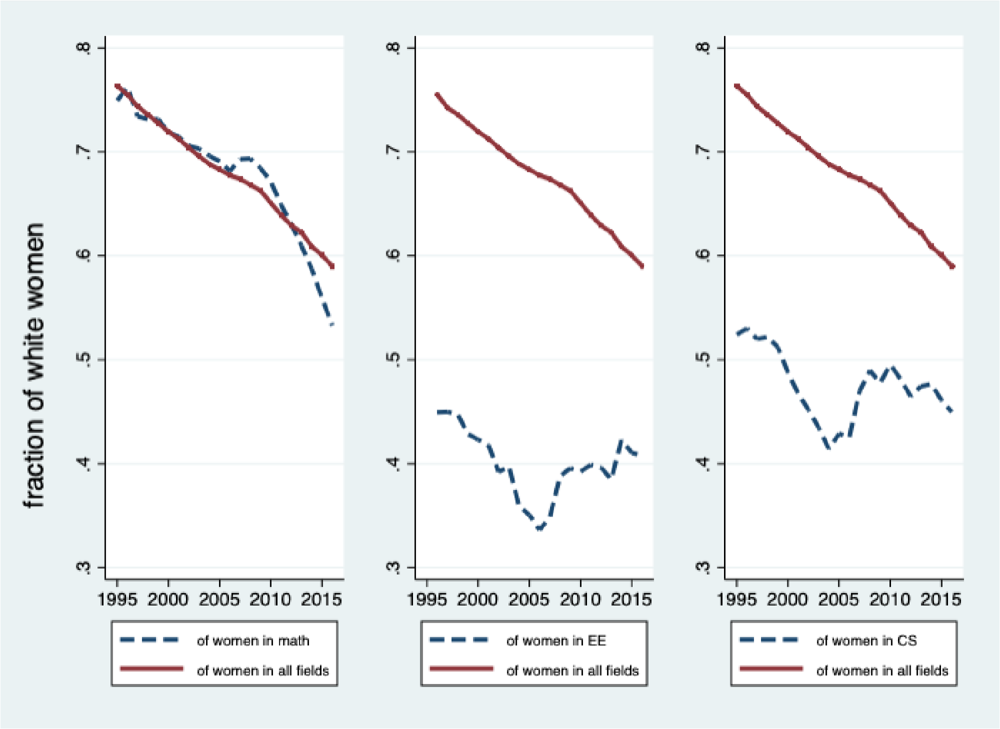
[13] These racial differences are not accounted for by differences in technical abilities, but rather likelihood in studying engineering.

Because CS is primarily associated with electrical engineering, white womens under-representation in electrical engineering is correlated with their under-representation in CS in particular. As figure 3.5 shows, white women's relative share in CS follows the same trend as their relative share in electrical engineering; Asian American women are over-represented in both electrical engineering and CS, but their trend in electrical engineering better matches with their relative presence in CS (figure C.4). The comparison of women from different races demands a "cultural" explanation. If there is a culture where it is normal for women to study or work in engineering, would there be also more women in computer science? Searching outside the U.S., there are some exceptions. In Malaysia, the information technology sector is made up equally of women and men. Before IT sector emerged, Malaysia women had been predominant in the Malay electronics industry (Huyer, 2015).

---

[13] The numbers are computed from the 2016 IPEDS survey (Integrated Postsecondary Education Data System, 1986–2016).

Figure 3.5: White women's representation in math, EE & CS



Notes: This figure depicts white women's relative representation in math, electrical engineering (EE) and computer science (CS). The solid lines plot the fraction of white women among female college graduates across all fields. The dashed lines plot the fraction of white women among female college graduates in math, EE or CS. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

3.3 Literature & Contribution

This paper relates to two strands of the literature. First, it contributes to the growing literature on the gender gap in STEM fields. By now, numerous studies have ruled out gender differences in math ability. If we look at test scores, the gender gap in mathematics achievement and aptitude is small and diminishing (Goldin and Kuziemko, 2006; Reuben et al., 2014; Xie and Shauman, 2003), and (Guiso et al., 2008; Stoet and Geary, 2018) suggest that cultural influences predict math achievement better than biology. Using cross country evidence, both studies show that in countries with higher gender equality, the performance gap between boys and girls in math is smaller. Paradoxically, these countries also have smaller shares of women in STEM (Stoet and

Geary, 2018).[14] Overall, math or science achievement at the primary and secondary levels poorly predicts whether a student majors in STEM in college. It cannot explain the higher relative likelihood that men choose to major in sciences and engineering (Turner and Bowen, 1999; Xie and Shauman, 2003), nor why mathematically talented women preferred careers in law, medicine, and biology over careers in physical sciences and engineering (Lubinski and Benbow (1992)).

In searching for other explanations, the literature has turned to discrimination. Women are discriminated against not based on their ability, but based on the stereotype that women are worse at technical or quantitative tasks. In a hiring experiment, Reuben et al. (2014) shows that both men and women are less likely to hire women for arithmetic tasks. The gap persists even when employers receive information about the candidates past performance. In an analogous experiment, professors assess male candidates as being more competent (Moss-Racusin et al., 2012). This stereotype likely impairs women's chances of obtaining a graduate degree in STEM and slows down their professional trajectory relative to men (Sarsons (Working Paper)).

Meanwhile, the under-representation of women in STEM may perpetuate itself through the low availability of female students and advisors. The presence of other women in male-dominated programs results in lower attrition rates for women, preventing exiting especially in the first year (Bostwick and Weinberg, 2018; Huntington-Klein and Rose, 2018). Female students with an advisor of the same gender tend to be more productive during the PhD and more likely to become professors themselves (Gaule and Piacentini, 2017). However, the results from these studies are threatened by external validity. Depending on where the sample comes from, say a different field, university or country, exposure to female peers or advisors have null or negative effects (Hilmer and Hilmer, 2007; Kunze and Miller, 2014; Neumark and Gardecki, 1998).

Because of the limited extent to which a student can try out different majors, choosing a college major is a decision made under uncertainty. Subjective expectations on financial as well as non-pecuniary outcomes related to each major calibrate each individual's academic preferences. The gender gap in undergraduate majors can be accounted by the systematic differences in male and

---

[14] The authors suggested that life-quality pressures in less gender-equal countries promote girls and womens engagement with STEM subjects.

female priorities. Wiswall and Zafar (2015) found that pecuniary outcomes in the workplace matter much more for male students whereas non-pecuniary outcomes, such as beliefs about enjoying coursework, enjoying work at potential jobs, and gaining parents' approval weigh more heavily in the decisions of female students. The finding that male students care more about economic returns is in keeping with the traditional gender division where men are seen as the breadwinners in marriage (Bertrand et al., 2015).

A byproduct of my research is a breakdown of STEM data STEM by discipline. While most studies concerning gender inequality look at STEM at the aggregate level, the variation in gender gaps across STEM subjects is too significant to ignore. In particular, there is a difference between science and propositional knowledge vs engineering and prescriptive knowledge, with the latter being much more male dominated.[15] This concept is well documented by historians and psychologists (Bix, 2014; Holth and Mellstrom, 2011). Few administrators would declare this in print today, [16] but the culture of masculinity and hierarchy is so deeply entrenched in engineering education that is evident to casual observers (de Pillis and de Pillis, 2008). My study builds on this subtle but pervasive gendering to explain why CS programs in engineering schools have a smaller proportion of women.

Historians have documented how and why women left computer science: programming started as a job for women. Hicks (2017) describes how computing experienced a gender flip in the United Kingdom and became male-identified by the civil service and public sectors in the 1960s and 1970s. Ensmenger (2010) confirms the feminine origin of programming in the U.S. experience, and argues that the difficulty in raising programming in professional status is tied to the job being seen as held by women. This turn of events, where a job goes from being predominantly held by women to mostly men, is rare.

Of course, no occupation is inherently male or female. The gendering of an occupation varies over locations, and evolves over time. As Goldin insightfully expressed: the "aura of gender" is

---

[15] Mokyr describes and differentiates these two kinds of knowledge and their roles in innovation in *The Gifts of Athena: Historical Origins of the Knowledge Economy* (Mokyr, 2002).

[16] Penn States engineering dean in 1955, stated that women are not suited to engineering, and that teaching them is wasted effort (Bix, 2004)).

gained through a rhetoric that surrounds the labor market, by the evolution of certain norms, and by the use of particular forms of physical capital (Goldin, 2002). My paper applies this insight to the case of computer science. I explore the changes in rhetoric surrounding this major as reflected in its departmental organization. When a CS department moves from the math department to the school of engineering, in effect, it becomes an engineering subject.

There is no single explanation for why the proportion of women in programming decreased. There are multiple plausible hypothesis including discrimination (Hicks, 2017), work climate (Chang, 2018), and the gendered marketing of personal computers (Henn, 2014). The reclassifying of CS to engineering is another explanation.

3.4 Data

The empirical design requires measurement in CS gender composition, CS departmental organizations and the characteristics of CS majors such as faculty, courses and degree requirements.

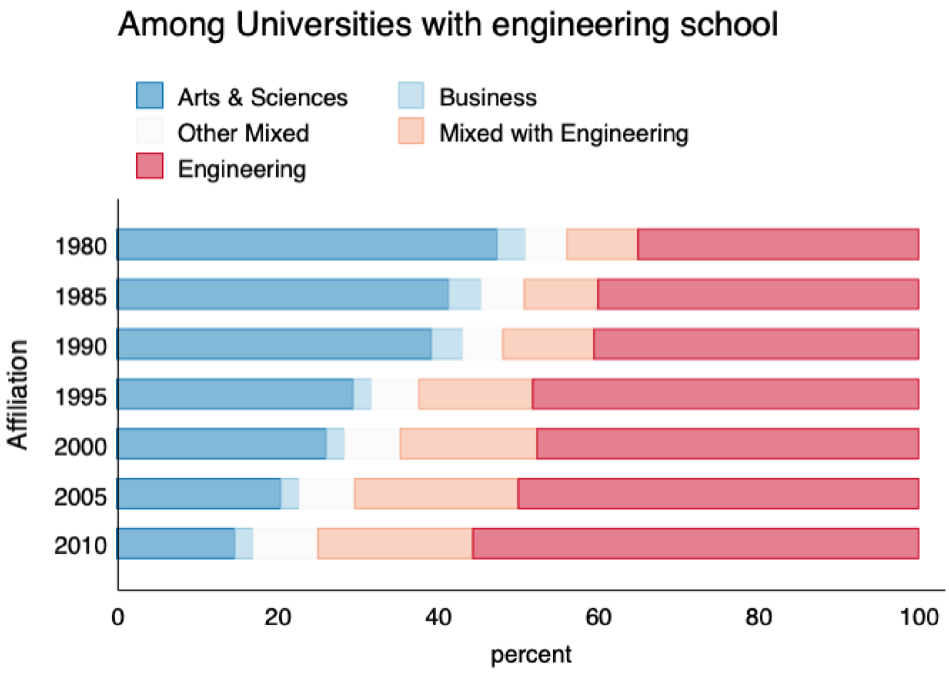
3.4.1 Department affiliation data

American universities have a high degree of freedom in how they organize their majors and minors, schools and departments. This information is published annually by each university in its course catalogs. I explore the heterogeneity in how universities organize their CS majors by taking advantage of this previously overlooked yet rich data source.

In this paper, I define departmental affiliation as the school within which the CS department is organized and administered. I refer to this type of affiliation as organization hierarchy in the rest of the paper. However, an affiliation can form in other ways, such as the sharing of faculty, sharing of courses and office buildings. I use the organization hierarchy as my measure of affiliation because this is the only measure consistently observed in all college and university catalogs.

The main source of data on organization hierarchy comes from the College Catalog Survey (CCS) (Brint, 2013). This is one of two databases produced by the Colleges & Universities 2000 Project. CCS includes 286 U.S. four-year colleges and universities, with measurements provided

Figure 3.6: Computer Science Departmental Affiliation



Notes: This figure shows the distribution of CS departmental affiliation from 1980 to 2010, among universities with engineering schools. Affiliation is defined as the schools that offer majors in CS. An affiliation is categorized as *Mixed with Engineering* if the CS major is offered in more than one school at a university, with one of the schools being the engineering school. An affiliation is categorized as *Other mixed* if the CS major is offered in more than one school at a university, but not in the engineering school. *Data source*: American College Catalog Study Database (Brint, 2013) at five-year intervals: 1975-6, 1980-1, 1985-6, 1990-1, 1995-6, 2000-1, 2005-6, and 2010-11. For each institution year observed, the name of the department and school are reported. CCS also contains coding for transitions, including splits of previously joined major units, moves of majors, consolidations of previously separate majors, reductions in status (e.g. from department to program status), and eliminations. In this dataset, I also observe when CS majors are offered through multiple schools.

Next, I categorize the CS affiliation for each university year observation into one of the following categories: arts (college of liberal arts & sciences), business (business school), engineering (engineering school), others (other uncategorized professional school), arts & business, arts & others, business & others, engineering & arts, engineering & arts & business, engineering & business, mixed within arts (two majors, both in college of liberal arts & sciences), mixed within engineering (two majors, both in engineering school), and mixed within others (two majors, both in uncategorized professional schools). Figure C.6 presents the distribution of CS affiliations over time among universities with engineering schools. Figure 3.6 presents a simplified version of figure C.6 by ignoring uncategorized professional schools and by dividing mixed cases into two subcategories: mixed with- and mixed without engineering. Both figures show that among universities with engineering programs, there is a decreasing share of CS programs affiliated with business schools or Colleges of Liberal Arts & Sciences, and increasing shares of CS programs with engineering-related affiliations.

Apart from organization hierarchy, course catalogs offer information on faculty, course offerings, degree requirements etc. I complement the CCS dataset by personally collecting the above data myself. For CS degree requirements, I count the maximum number of math, CS and engineering courses required to receive a bachelor's degree in minimum standards. Figure C.17 shows the distribution of math, CS and engineering courses required over time. While the total course load has remained stable since the 1980s, the number of math courses has decreased while the number of CS and engineering courses has increased.

3.4.2 Graduation data

Data on the CS gender ratio is constructed from the Higher Education General Information Survey (HEGIS) series and its successor, the Integrated Postsecondary Education Data System (IPEDS) series (Higher Education General Information Survey, 1965–1984; Integrated Postsecondary Education Data System, 1986–2016). HEGIS is available for the following years: 1965-1968, 1970-1972, 1974-1981 and 1983-1984. IPEDS is available from 1987-2016. Both surveys are conducted by the National Center for Education Statistics and contain data on earned degrees by subfield, degree level, sex and race.

Each subfield has an identifier code, which is the Item Number in HEGIS and the Classification of Instructional Program (CIP) in IPEDS. Since Item Number and CIP code classify subfields and the number of subfields increases over time, CIP coded additional and more granular subfields than captured by the Item Number. To harmonize the two series, I aggregate the Item Number and the CIP code to a broader field level. Taking Computer Science for example, I aggregate 5 Item Numbers: 2100 (computer science and systems analysis), 2105 (data processing), 2110 (computer science), 2120 (system analysis) and 2199 (computer science and systems analysis, unidentified) into the major "Computer Science" from the years 1965-1968. I aggregate Item Numbers 701-799, which contains 5 subfields from the years 1970-1972 & 1974-1981. I aggregate Item Numbers 110101-119999, which contain 6 subfields from the years 1983-1984. I aggregate CIP codes 11-11.9999 (42 subfields of computer science), 30.08 (mathematics and computer science) and 52.12 (5 subfields of business programming), into the CS major for the years 1987-2016.

To show that this way of harmonizing data makes sense, I present some continuous trends by using the broader field definition. Figure C.9 shows the harmonized trend in the fraction of women in CS as represented by the blue line. Figure C.11 shows the harmonized trends in the share of undergraduate women/men in CS. Further, to show that from the years 1987-2016, math & computer science and business programming should be included in the CS major, I plot their respective trends, which is presented in figure 3.10 (b). The fraction of women graduating in math & computer science and in business programming follow trends parallel to that of computer information system.

Degree completion data measures survival in a major program, which by its nature, measures the lagged outcome of any shock affecting a major. Since a relevant degree weighs significantly in the job search, the fraction of women completing a CS major *is* the final outcome that matters. Without measurement in immediate outcomes, I will be constrained in explaining or ruling out potential mechanisms that lead to the changes in degree completion. Ideally, I would also look at enrollment & grades in CS courses or the declaration of CS majors & GPA as more immediate outcomes.

To overcome this weakness, I exploit the variation already existing in the degree completion data, such as differences in response from lower and upper year students, and differences in response from undergraduate and master's students. I also bring together data from the American Community Surveyand an experiment to indirectly speak to the mechanism.

3.4.3 Some Illustrative Graphs

I merged graduation data with the affiliation data to create a panel of 243 colleges with CS programs. Since affiliation is observed at 5-year intervals, I averaged 3 years of graduation data around each set of 5 years.

Graphs for individual schools, although noisy, can provide intuition about the empirical strategy and the results. I consider three universities, chosen for their illustrative value. Indiana University at Bloomington has a CS department that has always been placed within the College of Liberal Arts and Sciences. This is the direct result of the lack of an engineering school. Northwestern University has a CS department that has always been located in the engineering school. In addition, it is affiliated with the department of electrical engineering. Clemson University founded their CS department in the college of liberal arts & sciences and relocated it to the College of Engineering between 1990 and 1995. Each of the figures 3.7 through 3.9 shows the fraction of undergraduate women graduating with CS, math and engineering degrees for the corresponding university, averaged over a three-year moving window. Figure 3.7 shows that-at Indiana University, the fraction of women in CS and math follow very similar trends. Figure 3.8 shows that-at Northwestern University, the fraction of women in CS is comparable to the fraction of women in electrical engineering. There seems to be no correlation with math or the engineering average. Figure 3.9 shows at Clemson University, the fraction of women in CS and math was comparable until Clemson moved the CS department to the engineering school, at which point the fraction of women dropped to the level of the engineering school. The two vertical red lines portray the period of time when the CS department moved.

These three examples illustrate two variations that I will explore in the analysis. Figures 3.7 & 3.8 explain the cross-sectional variation in CS affiliation. In addition, figure 3.8 suggests a meaningful variation within the engineering school. Electrical engineering is the hardware counterpart of CS. Even if they are not related in content, electrical engineering is the most male-dominated of all engineering subjects[17] and affiliating with electrical engineering could have an additional effect on the fraction of women.

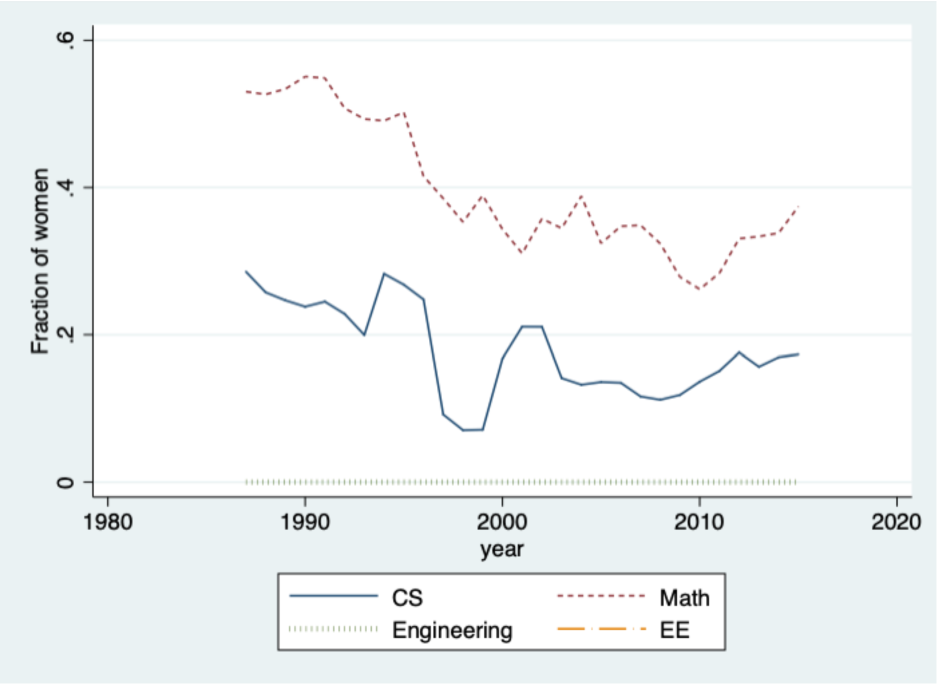
Figure 3.10 panel (a) shows the two variations in aggregate. It considers CS programs exclusively affiliated with engineering, liberal arts & sciences, business or other professional schools. It groups programs by affiliation and shows that cross sectionally, CS programs in engineering schools have a lower fraction of women compared to those in Business or Liberal Arts & Sciences.

Figure 3.11 panel (a) illustrates the within-university variation by dividing universities into three groups. The first group, the solid line, combines universities that never affiliated CS with engineering programs. The second group, the dashed line, combines universities that always affiliated CS with engineering programs. The third group, the dotted line, combines universities that at some point, re-affiliated CS programs with engineering. The fraction of women in CS from Groups 1 & 2 more or less follow the parallel trends, and the fraction of women in CS from Group 3 converges with that in Group 2 over time.

---

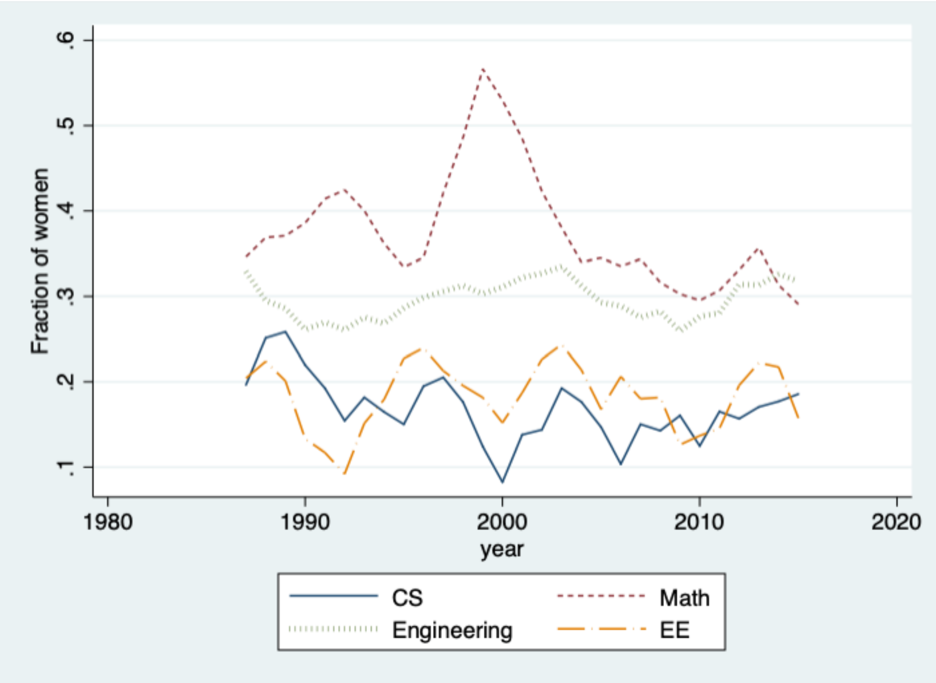
[17] see figure C.19

Figure 3.7: Indiana University trend



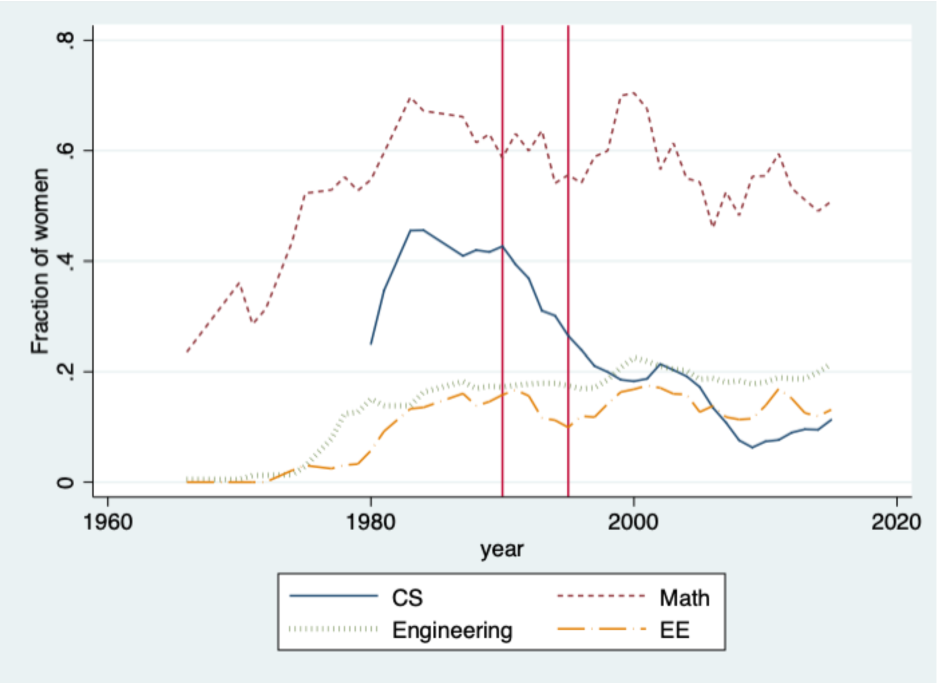
Notes: This figure plots the fraction of women receiving bachelor's degrees in Math and CS at Indiana University, averaged over a three year moving window. There is no engineering school at Indiana University. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

Figure 3.8: Northwestern University trend



Notes: This figure plots the fraction of women receiving bachelor's degrees in Math, CS, Engineering, and EE at Northwestern University, averaged over a three year moving window. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

Figure 3.9: Clemson University trend



Notes: This figure plots the fraction of women receiving bachelor's degrees in Math, CS, Engineering, and EE at Clemson University, averaged over a three year moving window. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

3.4.4 Comparison between the sample and the population

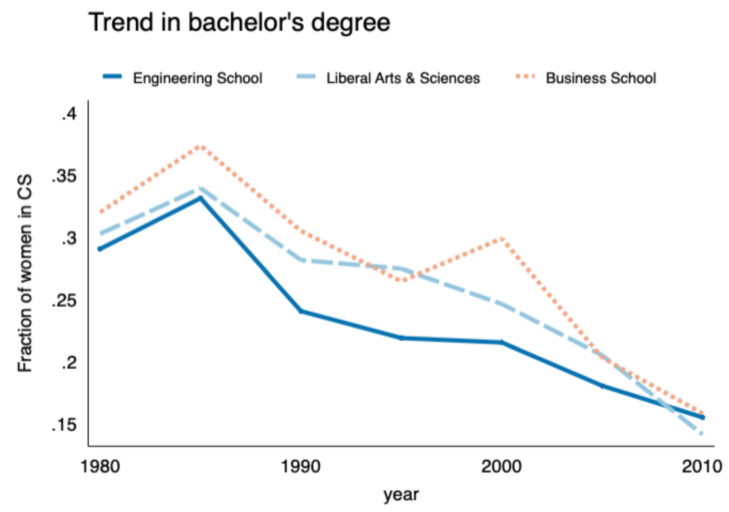
The sample consists of 286 institutions, of which 242 have undergraduate CS programs. That is roughly 12.9% percent of the population of 1,882 undergraduate CS programs offered in the U.S. The sample consists of colleges and universities that vary in sizes, private/public status, rank, and tier. Comparing to the population, the sample is weighted in favor of selective universities. Summary table 3.4 shows a comparison of the students and college characteristics in detail.

Even though selectivity might affect gender composition in a CS program, the sample and the population are comparable in their cross-section variation by affiliation. The sample suggests that CS programs in engineering schools have lower fractions of women than that in the college of liberal arts & sciences or the business school (figure 3.10 panel (a)). In the population, affiliation can be approximated by the degree CIP code. There are three groups of CIP codes relating to programming: computer information systems, business programming, and mathematics & computer science. While computer information systems may be located in both the college of liberal arts & sciences and the school of engineering, business programming is always located in the business school and mathematics & computer science is always located in the college of liberal arts & sciences. Consistent with the sample, computer information systema has a lower fraction of women compared to business programming or mathematics & computer science figure 3.10 panel (b).

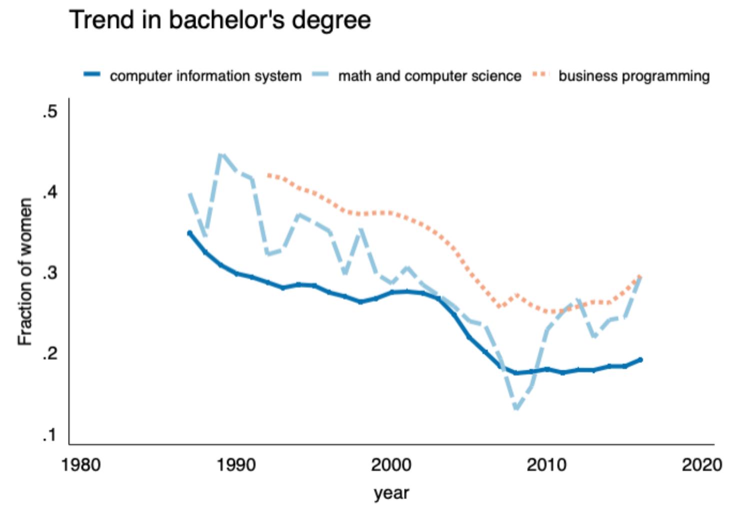
The sample and the population are also comparable in the variation over time. The sample groups CS programs based on whether they are located in engineering school: a group that always affiliates CS with engineering, a group that never affiliates CS with engineering and a switching group that changed between non-engineering and engineering affiliations. The fraction of women from the switching group converges with that from the group that always affiliates with engineering (figure 3.11 panel (a)). In the population, engineering affiliation can be approximated by whether a university has an engineering school. Analogously, universities in the population are binned into three groups. The first group combines universities that always have engineering programs. The second group combines universities that never have engineering programs. The third group combines universities that at some point had engineering programs. Just like in the sample, the fraction of women from the switching group is lower than that from the second group and converges with that from the first group (figure 3.11 panel (b)).

Figure 3.10: Fraction of women in CS by affiliation/subfield

(a) Sample: by affiliation



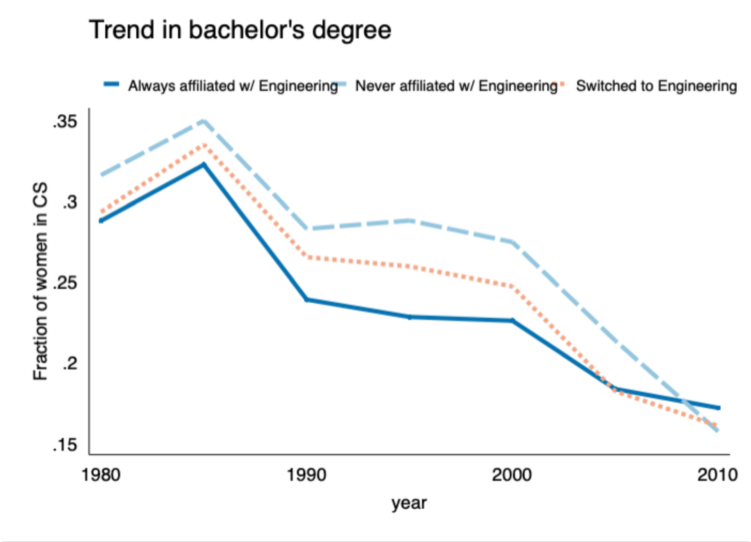
(b) Population: by subfield



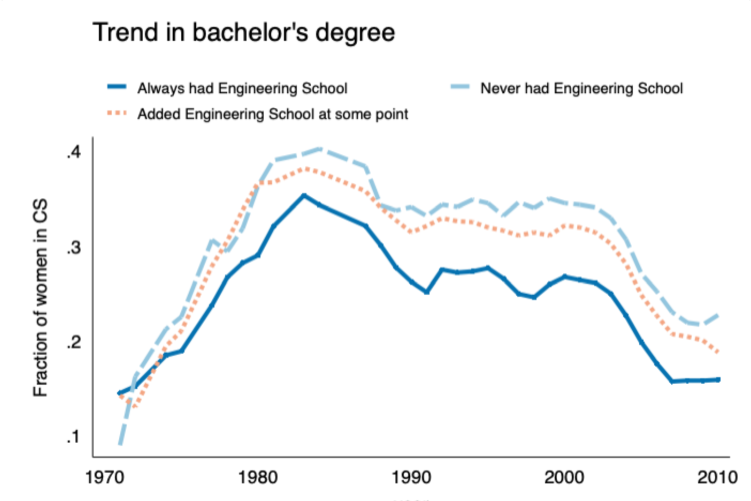
Notes: This figure visualizes the key cross sectional variations underlying the main analysis. Panel (a) uses colleges from Course Catalog Survey (Brint, 2013) and shows the average fraction of women in CS by affiliation. Panel (b) uses all colleges from IPEDS (Integrated Postsecondary Education Data System, 1986–2016) and shows the average fraction of women in CS by subfield.

Figure 3.11: Fraction of women in CS programs that “never”,“always”, or at “some” point–

(a) Sample: - were affiliated with engineering



(b) Population: - had an engineering program



Notes: This figure visualizes the key variation over time underlying the main analysis. Panel (a) uses colleges from Course Catalog Survey (Brint, 2013) and shows the average fraction of women in CS programs that never, always, or at some point affiliated with engineering. Panel (b) uses all colleges from IPEDS (Integrated Postsecondary Education Data System, 1986–2016) and shows the average fraction of women in CS programs that never, always and at some point locate in universities with undergraduate engineering programs.

3.4.5 Determinants of Engineering Affiliation

Before turning to the main empirical analysis, I document the patterns of how a university decides upon the school affiliation for CS. In particular, I examine whether the affiliation decisions were strategic, i.e., if they chose the affiliation to maximize the quality of their CS programs.

Using CS programs from universities with an engineering school, I first analyze whether universities with better-ranked engineering schools are more likely to associate CS with engineering. A better-ranked engineering school may provide more benefits to an engineering affiliation, through more funding resources, better facilities, and a better reputation to attract new faculty. In the analysis below, I define an engineering school as top-tier if its graduate program was ranked in the top 50 at least once by *U.S. News & Report* from 1980 to 2010.

Figure 3.12 presents the percentage of CS programs with engineering affiliation, grouped by whether the university has a top-tier engineering school. The top-tier group is consistently more likely to have an engineering affiliation from 1980 to 2010. Moreover, in 1980, even the top-tier group was evenly divided between an engineering affiliation and a non-engineering affiliation. I further explore the ranking within top-tier engineering schools and reveal that the universities with the best engineering schools switched first.

Table 3.3 shows the evidence on engineering ranking in logistic regressions. A unit of observation is a university that has both a CS program and an engineering school in the year 1980, 1990, 2000, or 2010. *engineering school Ranks Top 50* is a dummy variable capturing whether an engineering school is top-tier. *engineering school Ranking* is equal to the average ranking of an engineering school over the four years if it is a top-tier and equals to zero otherwise. The regressions include controls on general selectivity in undergraduate admission, size of enrollment, number of tenured faculty, and whether an electrical engineering program exists. Since 1990, the coefficient of *engineering school Ranking* shows that conditional on having a top-tier engineering school, a higher rank (i.e., a smaller number) increases the likelihood of a CS programs affiliating with engineering. In 2000, improving the ranking by 1-unit increases the odds of engineering affiliation by 5 percent.

The relationship between engineering school ranking and the predicted probability of CS having an engineering affiliation is visualized in figure 3.13. The logistic regressions include controls for Barron's competitive ranking, total fall enrollment, number of tenured faculty, and whether electrical engineering program is available. In 1980, the ranking of an engineering school had little relationship with engineering affiliation. By 1995, the relationship became significant, and a top 20 engineering school predicts an engineering affiliation with almost certainty. By 2010, the relationship disappeared again as the lower-ranked engineering schools caught up in affiliating with CS programs.

Top-tier engineering schools set the trend. They may have created the impression that those CS programs in engineering schools were better, encouraging other universities to follow. The CS program at Virginia Tech was a part of the College of Arts and Sciences for more than 30 years before it moved to the engineering school in 2004. The catalog offered the following explanation: "...the highest-ranked departments tend to function in an engineering rather than a sciences setting. In addition, the high-ranking and large-resource base of the College of Engineering should help support the department's strong, internally motivated efforts to improve as a department"(Virginia Tech, 2004).

Why were CS programs initially associated with mathematics departments? Practically, alliances with mathematics departments provided benefits for the creation of a CS curriculum. During the early stages of development, CS had few courses of its own and borrowed heavily from mathematics to compose its body of theory. Math courses not only appeared in CSs course offerings, but also made up a significant portion of the CS degree requirement. In my sample, an average CS degree in 1970 required 7 courses in CS and 5 courses in mathematics (figure C.17). At Stanford University for example, CS courses were first offered in the mathematics department even though Stanford had one of the best engineering schools in the nation. CS did not move to engineering until it had accumulated enough courses to be a department capable of offering its own major.

While the previous analyses examined only colleges that already had engineering schools, here I look at how the establishment of an engineering school affects the switch in affiliations for a CS program. Some CS programs are located in the college of liberal arts & sciences due to the lack of an engineering school at the university. Most liberal arts colleges-or universities that value liberal arts education simply do not offer professional degrees at the undergraduate level. Once an engineering school becomes available, affiliating CS with engineering becomes an option. For example, at UC-Riverside, CS moved to engineering to become part of the Computer Science and Engineering Department when the Bourns College of Engineering was established in 1990.

This option is not always productive, because not every engineering program provides the same positive spillover to CS. The synergy between CS and electrical or computer engineering is stronger than that between CS and civil, mechanical, and industrial engineering. Despite the lack of a shared curriculum at the undergraduate level, many of the faculty in CS had training or research interest in electrical or computer engineering and vice versa.[18] Therefore, the establishment of engineering schools that offer electrical or computer engineering should drive the re-affiliation of CS more than that of other engineering schools.

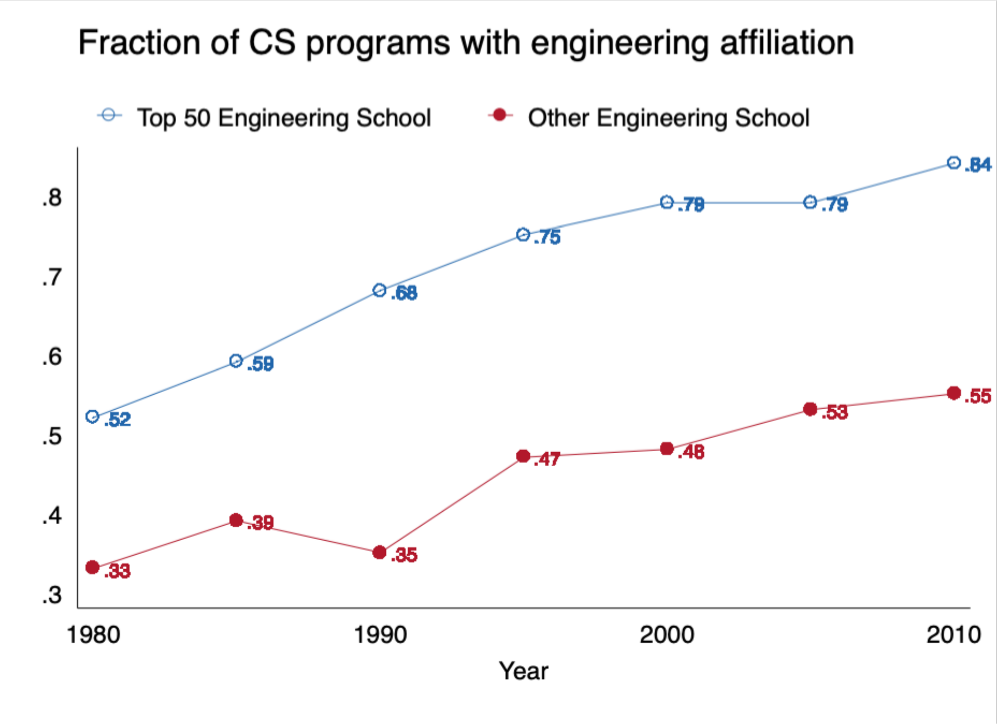
Table 3.1 presents the effect of establishing a engineering school and various engineering programs on a CS affiliation. Columns 1 & 2 from table 3.1 presents the estimates. If a university establishes an engineering school, the probability of it moving its CS program to the engineering school increases by approximately 20 percentage points. Columns 3 & 4 show that the relocation is driven by the opening of electrical or computer engineering programs rather than industrial, mechanical or civil engineering programs. This confirms that moving a CS program to engineering is driven by the incentive to affiliate with its hardware counterpart.

In summary, affiliation decisions are motivated by quality maximization. The best school to administer the CS program depends on what resources that school can offer at different stages of development.

---

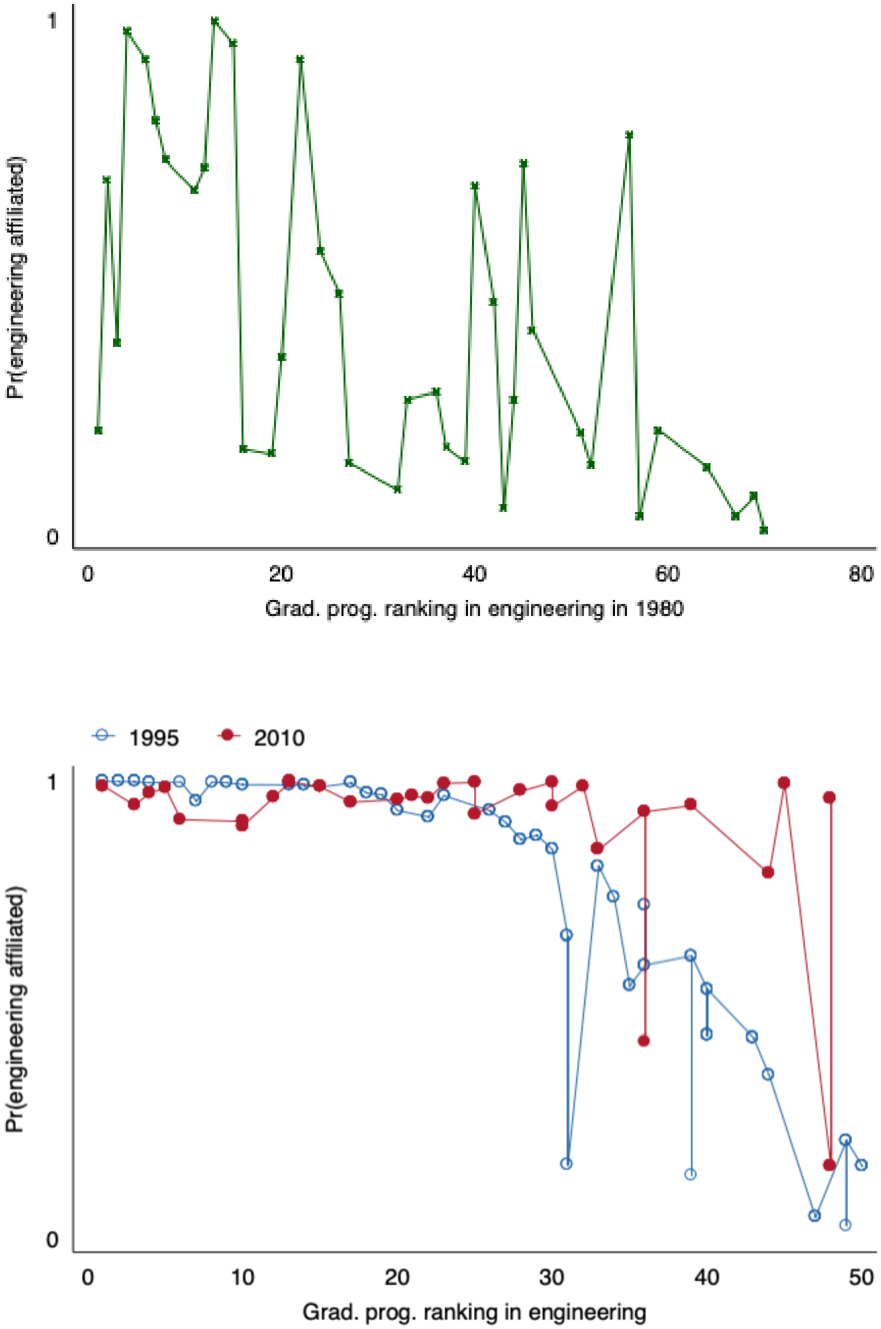
[18] For example, computer engineering designs chips for crunching neural network computation.

Figure 3.12: Top 50 engineering school as trendsetters



Notes: This figure presents the trends in the percentage of CS programs with an engineering affiliation, grouped by whether the residing university has a top 50 engineering school or a non top 50 engineering school. An engineering school qualifies as a top 50 if its graduate program ranked top 50 at least once by U.S. News & Report from 1980 to 2010. *Data source*: American College Catalog Study Database (Brint, 2013).

Figure 3.13: Engineering affiliation and graduate engineering program ranking



This figure displays the relationship between a university's ranking in engineering and the predicted probability that its CS program has an engineering affiliation. The controls in the logistic regression include Barron's competitive ranking, total fall enrollment, number of tenured faculty and whether EE program is available.

Table 3.1: Determinants of the switch

Engineering Affiliation

	(1)	(2)	(3)	(4)
	Exclusively Engineering	Including Engineering	Exclusively Engineering	Including Engineering
Engineering exists	0.2017***	0.2486***		
	(0.0369)	(0.0375)		
EE exists			0.2048***	0.1339**
			(0.0542)	(0.0566)
CE exists			0.1473***	0.1090***
			(0.0260)	(0.0272)
Industrial exists			-0.0932**	-0.0329
			(0.0405)	(0.0423)
Mechanical exists			0.0017	0.0040
			(0.0616)	(0.0644)
Civil exists			0.0155	-0.0110
			(0.0581)	(0.0608)
Year Fixed Effects	Yes	Yes	Yes	Yes
College Fixed Effects	Yes	Yes	Yes	Yes
Observations	1249	1249	1249	1249
R^2	0.815	0.840	0.823	0.838

Notes: The table reports OLS estimates. An observation is a college with a CS program(s) in year t. The dependent variable measures engineering affiliation either exclusively or inclusively. X exists is a dummy variable that equals 1 if the X program exists in year t. The regression controls for college fixed effects and year fixed effects. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

3.5 Main analysis and results

This section presents the empirical exploration of the relationship between engineering affiliation and the proportion of women in CS through OLS estimation. The estimates are identified by variation across time in a given college.

3.5.1 Panel Specification

The estimating equation is:

$$log(fraction\ of\ women_{c,t}) = \beta Engineering_{c,t} + \theta log(program\ size)_{c,t} + \lambda_c + T_t + \epsilon_{c,t} \quad (3.1)$$

where the outcome is the fraction of women in CS + 0.01 at college c in year t in logarithm, $Engineering_{c,t}$ is a dummy variable that equals 1 if CS affiliates with engineering at college c in year t, $log(program\ size)_{c,t}$ is the total number of students in the CS program at college c in year t in logarithm, λ_c is the college fixed effect, T_t is the year fixed effect, and $\epsilon_{c,t}$ is an error term.

The coefficient of interest, β, captures the relationship between engineering affiliation and the fraction of women in CS. A negative β provides evidence that the association with engineering, a male-dominated domain, turns CS into a male-dominated field.

Engineering affiliation is defined as either exclusive or inclusive. In the exclusive definition, the engineering school is the only school offering CS, whereas in the inclusive definition, the engineering school may offer CS alongside with the college of liberal arts & sciences, the business school or both. My sample identifies 326 engineering affiliations, of which 253 are exclusive.

Since program size is the denominator of the fraction, it must be controlled, as it correlates to the dependent variable by construction. A larger program is expected to have a positive impact on women. Given that a larger program is more likely to have the presence of female peers, it is easier to buffer against the stress from being in a male-dominated program. Moreover, CS training regularly involves group projects, so having female peers enables women to form study groups that include women. Table 3.2 shows that the coefficient on $log(program\ size)$ is indeed positive and significant: if program size increases by 10 percent, the fraction of women increases by 3 percent.

The inclusion of the college fixed effect allows me to flexibly address any time-invariant college characteristics that may differentially affect the tendency to incorporate CS as an engineering major. It controls for observables such as the location of the college, selectivity, public or private status, liberal arts or comprehensive curricula, etc. It also controls for unobservable cultural factors such as a college's heritage in professional education, attitude about interdisciplinary majors, etc.

The inclusion of the year fixed effect eliminates time trends that impact both re-affiliation decisions and gender composition in CS, such as the evolution in computer hardware and developments in the IT labor market. Controlling for year fixed effect allows me to employ the randomness in the timing of re-affiliation decisions.

Standard errors are clustered at the college-level to correct for their correlation within clusters (Bertrand et al., 2004) and observations are weighted by the number of graduates to capture the presence of heteroskedasticity with respect to program size (Solon et al., 2015).

Table 3.2 reports the main findings. Columns 1 & 2 in table 3.2 show the results of the baseline panel regression for exclusive and inclusive affiliation, respectively. Column 3 expands on the specification in column 2 to include interaction terms $engineeringXarts$ and $engineeringXbusiness$. Here, $engineering$ equals 1 if engineering is a inclusive affiliation. $engineeringXarts$ equals 1 if CS is offered in both the engineering school and the college of liberal arts and sciences. The interaction terms measure the marginal effects of making CS available through liberal arts or business, conditional on inclusive re-affiliation with engineering.[19] Column 4 expands on the specification in column 1 to include the interaction term $engineeringXEE$, where EE is a dummy variable that equals 1 if "electrical engineering" appears in the department title. This interaction term measures the marginal effect of EE affiliation conditional on exclusive re-affiliation with engineering.

The estimates are consistent with the graphical evidence in figure 3.11. Moving CS to the engineering school exclusively lowers the fraction of women by 15 percent. The negative impact can be alleviated by providing a larger program, or alternative programs within the college of liberal arts & sciences or the business school. It can be made worse by the additional tie with electrical engineering.

---

[19] CS programs that are simultaneously offered in different schools may vary in major title and degree requirement.

Table 3.2: Engineering Affiliation Effect

log fraction of Women in CS	(1) Exclusively Engineering	(2) Including Engineering	(3) Including Engineering	(4) Exclusively Engineering
engineering	-0.153***	-0.051	-0.122**	-0.092*
	(0.0502)	(0.0479)	(0.0520)	(0.0548)
log program size	0.318***	0.313***	0.322***	0.320***
	(0.0259)	(0.0261)	(0.0263)	(0.0258)
engineeringXarts			0.257***	
			(0.0851)	
engineeringXbusiness			0.181**	
			(0.0858)	
engineeringXEE				-0.262***
				(0.0979)
Year Fixed Effects	Yes	Yes	Yes	Yes
College Fixed Effects	Yes	Yes	Yes	Yes
Observations	631	631	631	631
R^2	0.559	0.553	0.562	0.566

Notes: The table reports OLS estimates. An observation is a college with a CS program(s) in year t. The dependent variable is the fraction of women in CS+0.01 in logarithm. *Engineering* is a dummy variable that equals 1 if the CS program is affiliated with engineering. Engineering affiliation could be either exclusive or inclusive. It is *exclusively engineering* if CS is only offered in the engineering school. It is *Including Engineering* if the engineering school is one of the affiliations. *EngineeringXarts* is an interaction term that equals 1 if a university offers CS program in both the school of engineering and the college of liberal arts & sciences. *EngineeringXbusiness* is an interaction term that equals 1 if a university offers CS programs in both the school of engineering and the college of business. *EngineeringXEE* equals 1 if a CS program belongs to the department of electrical engineering and computer science at the school of engineering. Log program size is the total number of CS graduates at college c in year t in logarithm. The regression controls for college fixed effects and year fixed effects. Standard errors are clustered at the college level and observations are weighted by the number of CS graduates. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

3.5.2 Discussion & Interpretation

In this section, I discuss the implications of the OLS estimates and interpret the findings to rule out competitive mechanisms. I argue that gender norms–how normative it is for women to be students in each discipline–drives the results.

If we measure how male-associated a discipline is by the percentage of women within it, then these gender norms explain the finding that more women are receiving CS degrees in the business school and in the college of liberal arts & sciences than in the engineering schools. Since electrical engineering is the most male-dominated of all engineering fields, this is also consistent with the finding that the additional affiliation with electrical engineering predicts a lower proportion of women in the student body than an affiliation with the other engineering programs.

However, one might not expect students who select into CS to follow the same gender norms regarding engineering. To show that this is not the case, I focus on a sample of CS graduates who earned a second degree. I utilize the American Community Survey to compute the distribution of the second major by sex and by year (figure C.20). As it shows, CS graduates do not deviate from gender norms: both men and women are as likely to double major CS with math and business; but until recently, only men would double major CS with engineering.[20] The second major also tracks what a CS graduate wants to do with his or her CS degree: women were equally interested in the theoretical aspects of CS and applications in the business world, but they did not see themselves becoming engineers.

The college fixed effects in the preceding section excluded differences in college fundamentals as competing explanations. However, even within a college, different departments and schools can vary in culture, resources, ranking, and requirements. I acknowledge that moving the CS program to the engineering school incurs more than a change in the degree of male association. There would also be changes in faculty, admissions, course content, and degree requirements that could affect gender composition in the CS major. In what follows, I will address several competing stories.

One of the competing stories is that the cost associated with earning a degree from the engineering school discourages women. Because women are more likely to start college in liberal arts, having to transfer to the engineering school imposes higher costs for women than for men. The added cost includes applying to the engineering school and satisfying the different general distribution requirements for engineering. However, my data disprove this story as the primary channel. As the interaction term in Column 4 table 3.2 shows, electrical engineering has a further negative impact on the proportion of women. Since electrical engineering has the same admission and the same general distribution requirements as other engineering majors, the costs to acquire prerequisites cannot explain the gender gap between CS in electrical engineering and that in other engineering majors.

Thinking more broadly, entry costs should play a small role, if any role at all, in womens choices of major. Half of the seats in medical schools and law schools are awarded to women [21].

---

[20] If we exclude Asian women, the share of women who double major with engineering does not change. See C.20 Panel (d).

[21] Sources on law school: ABA Required Disclosures (Standard 509 Reports), American Bar Association Section of Legal Education and Admissions to the Bar. Sources on medical schools: 2017 Applicant and Matriculant Data Tables, American Association of Medical Colleges.

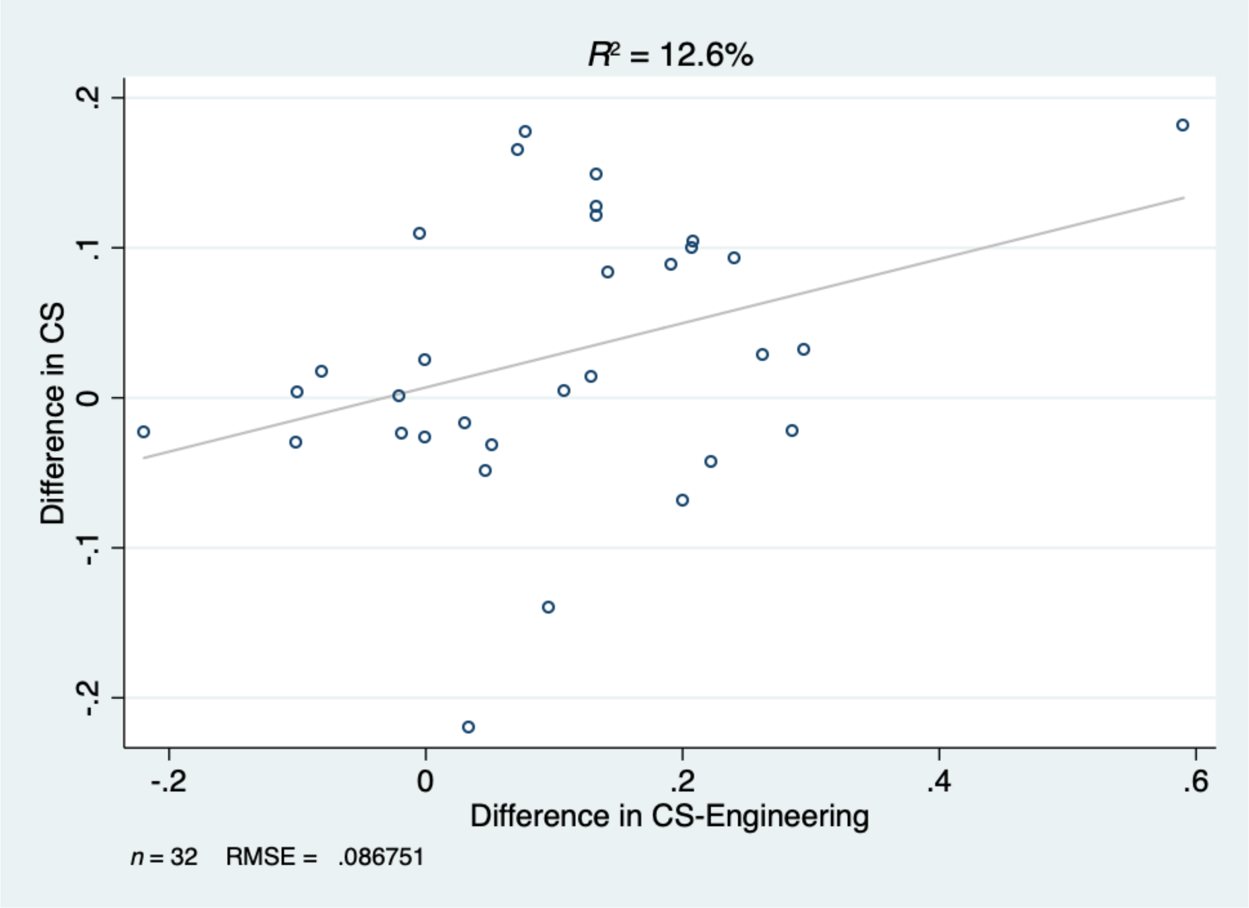
Both professional schools demand more competitive admissions than transferring to the engineering school within the same university. Moreover, business schools also have specific admission and general distribution requirements, yet CS programs in business schools have larger proportions of women.

An alternative explanation is that a masculine culture causes women to be unproductive in the engineering school. While a program's male-dominated student body is not a problem by itself, it is often a manifestation of masculine culture. Banchefsky and Park (2018) finds that male-dominated academic majors are more likely to endorse assimilation-that women should adapt and conform to masculine work norms in order to succeedand segregationthat men and women should pursue traditional social roles and careers. They were also more likely to agree with the gender-science stereotype that men do better in math and science than women. This masculine culture causes women to feel marginalized (Cheryan et al., 2017) and affects their academic performance (Walton et al., 2015). I confirm that this is a valid concern in the context of CS programs: in the cross-section, a CS program is more male-dominated if its affiliated engineering school is more male-dominated (figure C.15).

Moreover, by looking at CS programs that were about to move to engineering, I find that the drop in the female share in CS is positively correlated with the gap in the female share between CS and engineering at the time of the re-affiliation: the more male-dominated the engineering school is into which a CS program moved, the more the fraction of women in CS dropped. Figure 3.14 visualizes this correlation and Table 3.10 reports the estimates. The interaction term $engineeringXdifference$ measures the marginal effect of the gap conditional on engineering affiliation. Even though masculine culture played a significant role, neither the interaction term nor program size crowd out the estimate on engineering affiliation. This mechanism does not compete with my suggested mechanism.

Another major concern is that relocating the CS program to the engineering schools involves significantly more demanding coursework. This condition is hard to check. Using data collected from several degree requirements (figure C.16 shows the form of raw data), I find that the changes

Figure 3.14: CS Engineering gender gap & drop in CS's women share



in course load can go either way. In some cases, moving to the engineering school reduces the course load. Clemson University went down from 15 to 11 CS courses and 6 to 4 math courses required, Virginia Tech cut down 2 CS courses required, and SUNY buffalo also reduced CS course load from 15 to 11. In some other cases, the course load increased. Penn State added 3 more required CS courses, UC-Irvine increased CS course load from 19 to 25. Regardless of the changes in CS course load, relocating to the engineering school very rarely involves the addition of engineering courses. Except for Virginia Tech, which added two engineering courses, none of the schools mentioned above added any engineering courses to the degree requirement.

Even if there is no change in the curriculum, women may still believe that the courses in an engineering school will be more difficult. The best way to test this belief and disentangle the labeling effect from content is experimentation in a controlled setting. I designed a survey experiment where all the participants receive the same degree requirements and list of course offerings about a sample CS major. The only difference is that CS is framed as a science major in the control group and as an engineering major in the treatment group.[22] I then compare beliefs about CS majors and interests in pursuing a CS degree.

I ran a pilot experiment on 29 high school girls from less privileged neighborhoods in Chicago. I targeted this demographic because none of them had previous exposure to computer science in high school. Thus, they would be more likely to respond to framing. I find that the treatment group showed a lower level of interest in pursuing a CS degree. To test whether they expect that courses will be more difficult in the engineering school, I compare beliefs about completing the degree on time and completing with a minimum GPA of 3.0. I find no difference between the control group and the treatment group in these beliefs. What the engineering framing impact is the belief about how much they would enjoy the coursework. On a scale of 1-10, the treatment group scored 1.3 lower in their expectations about enjoying the coursework (Table C.2), and the belief in enjoying the coursework significantly predicts interest in pursuing a CS degree. In future work, I plan to find out whether gender norms explain the emotional response.

Finally, the negative relationship between the fraction of women and engineering affiliation only exists at the undergraduate level. There was no evidence that re-affiliation affected CS students from masters programs (Table 3.8). CS programs at master's level also have a higher percentage of women (30.4% in 2015). Because foreign male and female students constitute more than half of the male and female CS graduates from the masters programs, their dominance suggests that the gendered academic preferences towards math and engineering are specific to the U.S. context.

---

[22] I frame CS as a science major using a multiple-choice question: If you were to major in Science in college, what would be your most likely major? In the options followed, CS is one of them. I frame CS as an engineering major in the analogous way.

3.6 Conclusion

This paper provides evidence that traditional gender roles still influence women's college major choices today. Women invest in human capital in accordance with their female identity and based on the skills perceived to benefit household management and child-raising.

An evolution of historical events, such as the spread of home economics, redefined feminine subjects. As math and science were incorporated into domestic studies, the gender gap in these subjects diminished over time. Under the status quo, where women tend to study math or science more than engineering, narrating computer science as an engineering subject practically made it masculine. I find that moving the CS college major from liberal arts & sciences to engineering leads to a decreasing proportion of women in computer science.

By showing that contemporary variation in gendered college major choices can be traced back to historical events and that a college major can become male-dominated by being associated of a male-dominated discipline, this paper supports the idea that no college major is inherently male or female. The gendering of a major is shaped by history and norms, as well as psychology and behavior reflecting them.

Tables

Table 3.3: Determinant of Engineering Affiliation

Engineering Affiliation				
	1980	1990	2000	2010
Engineering School Ranks Top 50	0.4586	0.8238	2.1132	0.5976
	(1.2842)	(1.2162)	(1.4280)	(1.5281)
Engineering School Ranking	-0.0137	-0.0305	-0.0556*	-0.0542
	(0.0276)	(0.0297)	(0.0312)	(0.0357)
Barron's Competitiveness Ranking	-0.3476	-0.7639***	-0.6397**	-0.1864
	(0.3586)	(0.3107)	(0.2783)	(0.2866)
Total fall enrollment	Yes	Yes	Yes	Yes
Number of tenured faculty	Yes	Yes	Yes	Yes
EE exists	Yes	Yes	Yes	Yes
Observations	54	80	92	82

Notes: The table reports Logistic estimates. An observation is a university that offers CS major and has an engineering school. The dependent variable is a dummy variable that equals 1 if the CS major is affiliated with the engineering school in the corresponding year. Engineering Affiliation ranked Top 50 is a dummy variable that equals 1 if the engineering school ranked top 50 by U.S. News & Report in one of the years. engineering school Ranking equals to the average ranking over the years among top 50 programs. engineering school Ranking equals 0 if it is not a top 50 program. Barron's Competitiveness Ranking measures the admission selectivity of a university on the scale of 1-6, with 1 being the most competitive. Controls include total fall enrollment and number of tenured faculty. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table 3.4: Summary Statistics of students and college characteristics in 2010

	Population			Sample		
Panel A: Students	N	Mean	S.D.	N	Mean	S.D.
SAT Reading 75th percentile score	1016	585.156	68.99	195	620.877	79.19
SAT Math 75th percentile score	1028	594.414	70.07	196	632.505	82.18
Admission rate: enrollment/ applications						
Total	1364	0.238	0.12	219	0.205	0.11
Men	1341	0.245	0.13	215	0.211	0.11
Women	1361	0.231	0.12	219	0.202	0.11
Total full time enrollment	1364	988.869	1208.60	219	1691.849	1655.04
Panel B: Colleges						
Total full-time faculty	978	415.186	711.65	168	857.018	1046.29
In-state average tuition	1591	16799.085	10716.45	226	18361.283	14060.44
Out-of-state average tuition	1591	19734.731	8590.92	226	23820.288	10390.43
Share of public universities	1608	32.34%		226	50%	
Share of universities with engineering school	1608	31.09%		226	52.21%	

Notes: The table reports the mean, standard deviation, for students and colleges characteristics from the population and the sample. Each university had a CS program in 2010. *Data Source: IPEDS (Integrated Postsecondary Education Data System, 1986–2016).*

Table 3.6: Frequency of science terms mentioned in bibliography title

	science terms mentioned			
Subfields	biology	chemistry	psychology	public health
Applied Arts and Design	1	0	6	0
Child Care, Human Development	24	0	233	88
Clothing and Textiles	6	66	28	0
Food and Nutrition	9	35	11	15
Home and Institution management	0	1	4	2
Housekeeping and Etiquette	0	15	3	0
Housing, Furnishing and Home Equipment	0	1	4	8
Hygiene	47	16	5	305
Retail and Consumer studies	0	0	5	0
Total	87	134	299	418

Table 3.5: Home Economics bibliography: volumes and mean published year

Subfields	mean publish year	total volumes
Housekeeping and Etiquette	1890	1,231
Clothing and Textiles	1916	1,563
Hygiene	1916	1,088
Food and Nutrition	1917	1,274
Housing, Furnishing and Home Equipment	1917	706
Child Care, Human Development	1920	2,544
Home and Institution management	1920	676
Applied Arts and Design	1923	637
Retail and Consumer studies	1937	84

Table 3.7: Effect of offering Home Economics on women's participation in sciences

share of total women

	engineering	physical science	psychology	biology	chemistry	physics	geoscience	astronomy
Senior	0.0361	-0.0365*	0.0079	-0.0087	-0.0267	-0.0022	-0.0277	-0.0449
	(0.0388)	(0.0191)	(0.0190)	(0.0189)	(0.0182)	(0.0156)	(0.0264)	(0.0428)
Junior	0.0952**	-0.0779***	-0.0059	-0.0325*	-0.0735***	0.0056	-0.0275	-0.0381
	(0.0412)	(0.0192)	(0.0194)	(0.0193)	(0.0181)	(0.0162)	(0.0272)	(0.0463)
Sophomore	0.0264	-0.0400**	-0.0188	-0.0403**	-0.0426**	0.0121	-0.0183	0.0117
	(0.0399)	(0.0188)	(0.0190)	(0.0188)	(0.0178)	(0.0155)	(0.0256)	(0.0437)
Freshmen	-0.0237	-0.0401**	-0.0459***	-0.0415**	-0.0317**	0.0056	-0.0225	-0.0229
	(0.0343)	(0.0166)	(0.0168)	(0.0167)	(0.0156)	(0.0134)	(0.0229)	(0.0402)
Future cohort +1	-0.0057	-0.0374**	-0.0471***	0.0016	-0.0219	-0.0251*	-0.0284	-0.0322
	(0.0369)	(0.0179)	(0.0179)	(0.0178)	(0.0168)	(0.0148)	(0.0247)	(0.0428)
Future cohort +2	-0.0294	-0.0352*	-0.0467**	-0.0303	-0.0328*	-0.0075	0.0051	0.0100
	(0.0370)	(0.0186)	(0.0187)	(0.0187)	(0.0175)	(0.0150)	(0.0249)	(0.0414)
Future cohort +3	-0.0101	-0.0482***	-0.0159	-0.0314*	-0.0476***	-0.0164	-0.0366	-0.0060
	(0.0360)	(0.0179)	(0.0180)	(0.0181)	(0.0169)	(0.0150)	(0.0244)	(0.0422)
log total women	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
log total students	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
college fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
year trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17504	45539	51686	52582	43562	27947	16597	3488
R^2	.153022	.021595	.120365	.112349	.02012	.01193	.011492	.11895

Table 3.8: Engineering Affiliation Effect on Master's

log fraction of Women in CS

	Exclusively Engineering	Including Engineering	Including Engineering	Exclusively Engineering
engineering	0.127*	0.075	0.121	0.157*
	(0.0743)	(0.0746)	(0.0801)	(0.0817)
engineeringXarts			-0.135	
			(0.1234)	
engineeringXbusiness			-0.189	
			(0.1402)	
engineeringXEE				-0.104
				(0.1206)
log program size	-0.014	-0.013	-0.014	-0.015
	(0.0184)	(0.0184)	(0.0185)	(0.0185)
Year Fixed Effects	Yes	Yes	Yes	Yes
College Fixed Effects	Yes	Yes	Yes	Yes
Observations	481	481	481	481
R^2	0.461	0.459	0.462	0.462

Table 3.10: CS-Engineering Gap effect

log fraction of Women in CS	Including Engineering	Including Engineering	Exclusively Engineering	Exclusively Engineering
engineering	-0.0512	-0.1155*	-0.1621***	-0.1027
	(0.0598)	(0.0628)	(0.0583)	(0.0641)
engineeringXdifference	-0.0661*	-0.0773*	-0.0711*	-0.0713*
	(0.0396)	(0.0395)	(0.0398)	(0.0396)
engineeringXarts		0.2579**		
		(0.1060)		
engineeringXbusiness		0.2126**		
		(0.1047)		
engineeringXEE				-0.2324**
				(0.1056)
log program size	0.3552***	0.3594***	0.3611***	0.3638***
	(0.0378)	(0.0378)	(0.0374)	(0.0372)
Year Fixed Effects	Yes	Yes	Yes	Yes
College Fixed Effects	Yes	Yes	Yes	Yes
Observations	574	574	574	574
R^2	0.71	0.72	0.72	0.7

Table 3.9: Effect of opening engineering program on women's share

Share of Women in CS					
Senior	-0.0613***	-0.0194	0.0105	-0.0141	-0.0108
	(0.0147)	(0.0156)	(0.0173)	(0.0152)	(0.0111)
Junior	-0.0856***	-0.0208	0.0275	-0.0087	0.0089
	(0.0152)	(0.0160)	(0.0183)	(0.0163)	(0.0120)
Sophomore	-0.0340*	-0.0173	-0.0025	0.0188	0.0044
	(0.0194)	(0.0185)	(0.0184)	(0.0163)	(0.0126)
Freshman	-0.0521***	-0.0456***	-0.0060	-0.0264*	-0.0207**
	(0.0121)	(0.0134)	(0.0165)	(0.0139)	(0.0096)
Institution level controls	Yes	Yes	Yes	Yes	Yes
Institution fixed effect	Yes	Yes	Yes	Yes	Yes
State year effect trend	No	Yes	Yes	Yes	Yes
Year fixed effect	No	No	Yes	Yes	Yes
Program size $\geq$ 25	Yes	No	No	Yes	No
Year greater $\geq$ 1970	No	No	Yes	Yes	No
Weights on program size	No	No	No	No	Yes
Observations	11,719	24,068	23,911	11,308	24,068
R^2	.018	.13	.16	.34	.29

Notes: The dependent variable is the share of women in Computer Science. Observations are at the institution year level and the time units are year. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

References

ALBERT R. MANN LIBRARY (2005). Home Economics Archive: Research, Tradition and History (HEARTH). URL `http://hearth.library.cornell.edu`.

EDSGER W. DIJKSTRA (1978). The pragmatic engineer versus the scientific designer. URL `http://www.cs.utexas.edu/users/EWD/ewd06xx/EWD690.PDF`.

VIRGINIA TECH (2004). 2004-05 Undergraduate Course Catalog & Academic Policies.

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**.

ALESINA, A., GIULIANO, P. and NUNN, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, **128** 469–530.

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

ANDREWS, M. J. (2019). Local effects of land grant colleges on agricultural innovation and output. Working Paper 26235, National Bureau of Economic Research.

ATACK, J. (2013). On the use of geographic information systems in economic history: The american transportation revolution revisited. *The Journal of Economic History*, **73** 313338.

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

BANCHEFSKY, S. and PARK, B. (2018). Negative Gender Ideologies and Gender-Science Stereotypes Are More Pervasive in Male-Dominated Academic Disciplines. *Social Sciences*, **7** 27.

BECKER, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, **3** S33–S58.

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

BECKER, S. O., PFAFF, S. and RUBIN, J. (2016). Causes and consequences of the protestant reformation. *Explorations in Economic History*, **62** 1–25.

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.

BENSON, L. and SILBEY, J. (2006). Electoral and demographic data for new york,1830-1875. URL `https://doi.org/10.3886/ICPSR06926.v1`.

BERTRAND, M., DUFLO, E. and MULLAINATHAN, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, **119** 249–275.

BERTRAND, M., KAMENICA, E. and PAN, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, **130** 571–614. URL `https://doi.org/10.1093/qje/qjv001`.

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.

BIX, A. S. (2002). Equipped for life: Gendered technical training and consumerism in home economics, 1920-1980. *Technology and Culture*, **43** 728–754.

BIX, A. S. (2004). From “Engineeresse” to “Girl Engineers” to “Good Engineers”: A History of Women’s U.S. Engineering Education. *NWSA Journal*, **16** 27–49.

BIX, A. S. (2014). *Girls Coming to Tech! A History of American Engineering Education for Women*. The Mit Press, Cambridge, MA.

Blackwell, E. (2017). The domestic arts and sciences. *After Graduation Weinberg Magazine*. URL `https://www.weinberg.northwestern.edu/after-graduation/weinberg-magazine/fall-winter-2017/the-domestic-arts-and-sciences.html`.

Bostwick, V. K. and Weinberg, B. A. (2018). Nevertheless she persisted? gender peer effects in doctoral stem programs. Working Paper 25028, National Bureau of Economic Research. URL `http://www.nber.org/papers/w25028`.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, **90** 443–450. URL `http://www.jstor.org/stable/2291055`.

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

Brint, S. (2013). American college catalog study database, 1975-2011. URL `https://doi.org/10.3886/ICPSR34851.v1`.

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.

Bursztyn, L., Fujiwara, T. and Pallais, A. (2017). “Acting Wife”: Marriage Market Incentives and Labor Market Investments. *American Economic Review*, **107** 3288–3319. URL `http://www.aeaweb.orgarticles?id=10.1257/aer.20170029`.

CAG, 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, 60525.

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.

CECI, S. J., GINTHER, D. K., KAHN, S. and WILLIAMS1, W. M. (2014). Women in Academic Science: A Changing Landscape. *Psychological Science in the Public Interest*, **15** 75–141.

CHAN, R. Y. (2016). Studying philanthropy and fundraising in the field of higher education: A proposed conceptual model. In *Facilitating higher education growth through fundraising and philanthropy*. IGI Global, 1–27.

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

CHANG, E. (2018). *Brotopia: Breaking up the boys' club of Silicon Valley*. Portfolio/Penguin, New York, NY.

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

CHERYAN, S., ZIEGLER, S. A., MONTOYA, A. K. and JIANG, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, **143** 1–35.

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

CLAUDIA GOLDIN, L. F. K. (1999). Human capital and social capital: The rise of secondary schooling in america, 1910 to 1940. *Journal of Interdisciplinary History*, **29** 683–723.

COOK, L. D., LOGAN, T. D. and PARMAN, J. M. (2014). Distinctively black names in the American past. *Explorations in Economic History*, **53** 64–82.

COWAN, R. S. (1985). *More Work for mother: The Ironies Of Household Technology From The Open Hearth To The Microwave*. Basic books, New York.

DAN A. BLACK, S. G. S., AMELIA M. HAVILAND and TAYLOR, L. J. (2008). Gender wage disparities among the highly educated. *The Journal of Human Resources*, **43** 630–659.

DE PILLIS, E. and DE PILLIS, L. (2008). Are engineering schools masculine and authoritarian? the mission statements say yes. *Journal of Diversity in Higher Education*, **1** 33–34.

DIPPEL, C., GOLD, R., HEBLICH, S. and PINTO, R. (2019). Mediation analysis in iv settings with a single instrument. *Working Paper*.

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

ECCLES, J. S. and JACOBS, J. E. (1986). Social forces shape math attitudes and performance. *Signs*, **11** 367–380.

ELIAS, M. J. (2008). *Stir it up: Home Economics in American Culture*. University of Pennsylvania Press, Philadelphia, Pennsylvania.

ENGST, E. and FRIEDLANDER, B. (2014). Cornell rewind: The influence of eleanor roosevelt. *Cornell Chronicle*.

ENSMENGER, N. L. (2010). *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*. The MIT Press.

EPPRIGHT, E. S. and FERGUSON, E. S. (1971). *A century of home economics at Iowa State University. A proud past, a lively present, a future promise.* Iowa State University Press, Ames.

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.

FISMAN, R., IYENGAR, S. S., KAMENICA, E. and SIMONSON, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, **121** 673–697. URL `https://doi.org/10.1162/qjec.2006.121.2.673`.

FOLKE, O. and RICKNE, J. (3002forthcoming). All the Single Ladies: Job Promotions and the Durability of Marriage. *American Economic Journal: Applied Economics*.

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

GAULE, P. and PIACENTINI, M. (2017). An advisor like me? advisor gender and post- graduate careers in science. Working paper. URL `http://www.cerge-ei.cz/pdf/wp/Wp594.pdf`.

GEIGER, R. (1997). Research, graduate education, and the ecology of american universities: an interpretive history. chapter from graduate education in the united states.

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. (2002). A pollution theory of discrimination: Male and female differences in occupations and earnings. Working Paper 8985, National Bureau of Economic Research. URL `http://www.nber.org/papers/w8985`.

GOLDIN, C. (2006). The quiet revolution that transformed womens employment, education, and family.

GOLDIN, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104** 1091–1119.

GOLDIN, C. and KATZ, L. F. (1999). The shaping of higher education: The formative years in the united states, 1890-1940. *Journal of Economic Perspectives*, **13**.

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.

GOLDIN, L. K., C. and KUZIEMKO, I. (2006). The homecoming of american college women: The reversal of the college gender gap. **20** 133–156.

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

GRAY, R. (2013). Taking technology to task: The skill content of technological change in early twentieth century united states. *Explorations in Economic History*, **50** 351–367.

GUISO, L., MONTE, F., SAPIENZA, P. and ZINGALES, L. (2008). Culture, gender, and math. *Science*, **320** 1164–1165. `http://science.sciencemag.org/content/320/5880/1164.full.pdf`, URL `http://science.sciencemag.org/content/320/5880/1164`.

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**.

Henn, S. (2014). When women stopped coding. *National Public Radio*. URL `https://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding`.

Hicks, M. (2017). *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. 1st ed. The MIT Press.

Higher Education General Information Survey (1965–1984). Earned Degrees.

Hilmer, C. and Hilmer, M. (2007). Women helping women, men helping women? same-gender mentoring, initial job placements, and early career publishing success for economics phds. *American Economic Review*, **97** 422–426. URL `http://www.aeaweb.org/articles?id=10.1257/aer.97.2.422`.

Holth, L. and Mellstrom, U. (2011). Revisiting engineering, masculinity and technology studies: Old structures with new openings. *International Journal of Gender, Science and Technology*, **3** 313–329. URL `http://genderandset.open.ac.uk/index.php/genderandset/article/view/134`.

Hornbeck, R. A. (2010). Barbed wire: Property rights and agricultural developmen. *Quarterly Journal of Economics*, **125** 767–810.

Huntington-Klein, N. and Rose, E. (2018). Gender peer effects in a predominantly male environment: Evidence from west point. *AEA Papers and Proceedings*, **108** 392–95. URL `http://www.aeaweb.org/articles?id=10.1257/pandp.20181114`.

Huyer, S. (2015). Is the gender gap narrowing in science and engineering? *UNESCO SCIENCE REPORT*.

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.

INTEGRATED POSTSECONDARY EDUCATION DATA SYSTEM (1986–2016). Earned Degrees. URL `https://nces.ed.gov/ipeds/`.

INTERNATIONAL BUSINESS MACHINES CORPORATION (1974). IBM System/370 Principles of Operation. Manual. URL `http://bitsavers.informatik.uni-stuttgart.de/pdf/ibm/370/princOps/GA22-7000-4_370_Principles_Of_Operation_Sep75.pdf`.

IYER, S. (2016). The new economics of religion. *Journal of Economic Literature*, **54** 395–441.

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.

JONES, B. F. (2009). The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder? *The Review of Economic Studies*, **76** 283–317.

KAHN, S. and GINTHER, D. (2017). Women and stem. Working Paper 23525, National Bureau of Economic Research.

KUNZE, A. and MILLER, A. R. (2014). Women helping women? evidence from private sector data on workplace hierarchies. Working Paper 20761, National Bureau of Economic Research. URL `http://www.nber.org/papers/w20761`.

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.

LABAREE, D. F. (2017). *A perfect mess: The unlikely ascendancy of American higher education*. University of Chicago Press.

LUBINSKI, D. and BENBOW, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math-science pipeline. *Current Directions in Psychological Science*, **1** 61–66.

MANUELLI, R. E. and SESHADRI, A. (2014). Frictionless technology diffusion: The case of tractors. *American Economic Review*, **104** 1368–91. URL `https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1368`.

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.

MCCOLLOUGH, E. G. (1912). The cooperation of dietitian and physician. *Journal of Home Economics*, **4** 428–433.

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

MOKYR, J. (2000). Why ”more work for mother?” knowledge and household behavior, 1870-1945. *The Journal of Economic History*, **60** 1–41.

MOKYR, J. (2002). *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.

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.

MOSS-RACUSIN, C. A., DOVIDIO, J. F., BRESCOLL, V. L., GRAHAM, M. J. and HANDELSMAN, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, **109** 16474–16479. URL `http://www.pnas.org/content/109/41/16474`.

NEUMARK, D. and GARDECKI, R. (1998). Women helping women? role model and mentoring effects on female ph.d. students in economics. *The Journal of Human Resources*, **33** 220–246.

NYHART, L. K. (1997). Home economists in the hospital, 1900-1930. In *Rethinking Home Economics: women and the history of a profession* (S. Stage and V. B. Vincenti, eds.), chap. 6. Cornell University Press, Ithaca and London, 125–144.

OSTER, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, **37** 187–204.

PARNAS, D. L. (1998). Software engineering programmes are not computer science programmes. *Annals of Software Engineering*, **6** 19–37.

PERLMAN, E. R. (????). Tools for harmonizing county boundaries. URL `http://elisabethperlman.net/code.html`.

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

PRIESER, C. (1985). Occupational salary levels for white-collar workers, 1985 . Tech. rep., Bureau of Labor Statistics. URL `https://www.bls.gov/opub/mlr/1985/10/rpt1full.pdf`.

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.

RAMEY, V. A. and FRANCIS, N. (2009). A century of work and leisure. *American Economic Journal: Macroeconomics*, **2** 189–224.

RAYL, A. (2008). NASA Engineers and Scientists-Transforming Dreams Into Reality. *NASA 50th Anniversary Magazine*. URL `https://www.nasa.gov/50th/50th_magazine/scientists.html`.

REUBEN, E., SAPIENZA, P. and ZINGALES, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, **111** 4403–4408. `http://www.pnas.org/content/111/12/4403.full.pdf`, URL `http://www.pnas.org/content/111/12/4403`.

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

RUSH, B. (1787). Thoughts upon female education.

R.YOST, J. (2017). *Making IT work: a history of the computer services industry*. MIT Press, Cambridge, MA.

SARSONS, H. (Working Paper). Interpreting signals in the labor market: Evidence from medical referrals [job market paper].

SCHULTZ, T. W. (1975). The Value of the Ability to Deal with Disequilibria. *Journal of Economic Literature*, **13** 827–846.

SCHWIEDER, D. (1986). Education and change in the lives of lowa farm women, 1900-1940. *Agricultural History*, **60** 200–215.

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

SMITH, S. L. (1933). Research projects in home economics at the agricultural experiment stations and land grant colleges and universities 1931-1932. Tech. rep., U.S. Department of Agriculture.

SOLON, G., HAIDER, S. J. and WOOLDRIDGE, J. M. (2015). What Are We Weighting For? *Journal of Human Resources*, **50** 301316.

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

SPANGENBURG, R. and MOSER, D. (2003). *African Americans in science, math, and invention*. Facts On File, New York, NY.

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

STAGE, S. and VINCENTI, V. B. (1997). *Rethinking Home Economics: Women and the History of a Profession*. Cornell University Press.

STEVEN RUGGLES, R. G. J. G. E. M. J. P., SARAH FLOOD and SOBEK, M. (2019). IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS. URL `https://doi.org/10.18128/D010.V9.0`.

STOET, G. and GEARY, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, **29** 581–593. PMID: 29442575, URL `https://doi.org/10.1177/0956797617741719`.

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.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, **58** 267–288.

Turner, S. and Bowen, W. G. (1999). Choice of major: The changing (unchanging) gender gap. *ILR Review*, **52** 289–313.

U.S. Dept. of Labor, Bureau of Labor Statistics (1976). Occupational projections and training data .

U.S. Dept. of Labor, Bureau of Labor Statistics (2000). Occupational projections and training data .

Vidart, D. (2020). Human capital, female employment, and electricity: Evidence from the early 20th century united states. Job market paper.

Walton, G. M., Logel, C., Peach, J. M. and Spencer, M. P., Steven J. andZanna (2015). Two brief interventions to mitigate a chilly climate transform womens experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, **107** 468–485.

Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, **82** 791–824.

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

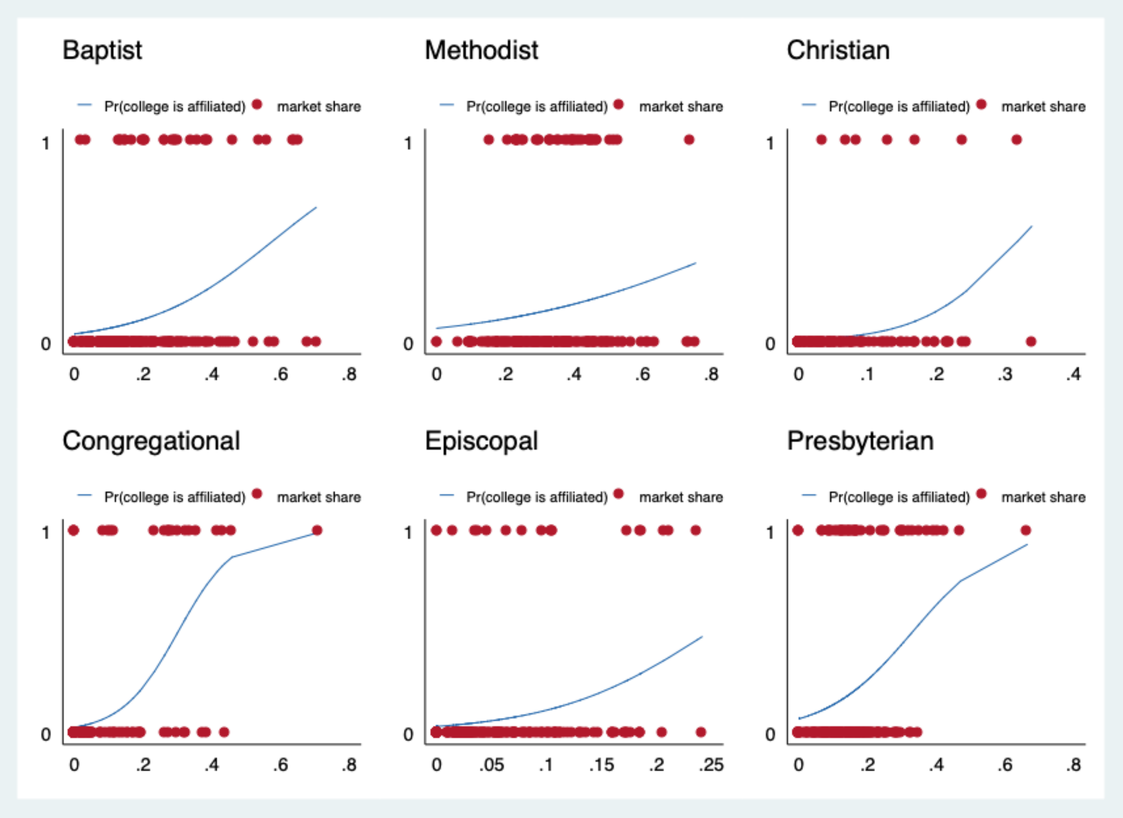
Xie, Y. and Shauman, K. (2003). *Women in Science: Career Processes and Outcomes*. Harvard University Press, Cambridge, MA.

Appendices

APPENDIX A

CHAPTER1

Figure A.1: Denomination Market share & affiliated college



Notes: each figure shows the predicted probability that the college is affiliated with a certain denomination based on the county-level market share of that denomination. Data are from the 1860 sample.

APPENDIX B

CHAPTER2

Table B.1: Spill over to science: full sample

Women in Science	Full Sample	Full Sample	Full Sample	Land grant Sample
Women in HomeEc	0.0444**			0.0928***
	(0.0211)			(0.0220)
Women in Classics		-0.0019		-0.0090
		(0.0068)		(0.0130)
Women in Education			-0.0020	-0.0181
			(0.0098)	(0.0260)
Land grant status	Y	Y	Y	N
public funds	Y	Y	Y	Y
private funds	Y	Y	Y	Y
Size	Y	Y	Y	Y
Men in Science	Y	Y	Y	Y
Female College	Y	Y	Y	Y
State fixed effects	Y	Y	Y	N
Observations	578	578	578	42
R^2	.1	.1	.1	.42

Notes: The table reports OLS estimates. An observation is a college in 1910. The dependent variable is the number of degrees in general science conferred to women. Women in home ec/classics is the number of degrees in general science conferred to women in home economics/classics. Number of majors is the total number of majors offered at the college. Scientific apparatus measures the value of scientific apparatus and teaching expenses on sciences measures the disbursement of funds on the teaching of natural sciences. Men in science equals to the number of degrees in general science conferred to men. Total faculty/books/endowment are measured in logarithm. Funds from public sources equals the amount of funds from state, federal government in logarithm. Funds from private sourses is measured in logarithm. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

Table B.2: Cp, R-squared and Actions along the sequence of Lasso Algorithm

Step	Cp	R^2	Action
1	74.8429	0	
2	41.6181	0.2891	+value of farm implements/machinery
3	11.7252	0.5509	+state population
4	2.4243 *	0.6436	+total enrollment at landgrant
5	3.3911	0.6521	+improved acres
6	4.7873	0.657	+agriculture population
7	5.3482	0.6689	-agriculture population +no. farms
8	5.5579	0.6835	
9	7.0676	0.6876	+value of livstock
10	9.0135	0.688	+value of farmland
11	11	0.6881	+agriculture population

Notes: This table reports the lasso solution in a linear regression where the dependent variable is total enrollment in agriculture. The computation algorithm follows a modified least angle regression (Efron et al., 2004), and the Stata code is implemented by Mander (2006). An observation is a land-grant university in 1910. The lasso algorithm selects among land-grant enrollment and the following agricultural variables in the corresponding state: value of farm implements/machinery, state population, improved acres, population in agriculture, number of farms, value of livestock, and value of farmland. * indicates the smallest value for Cp statistic.

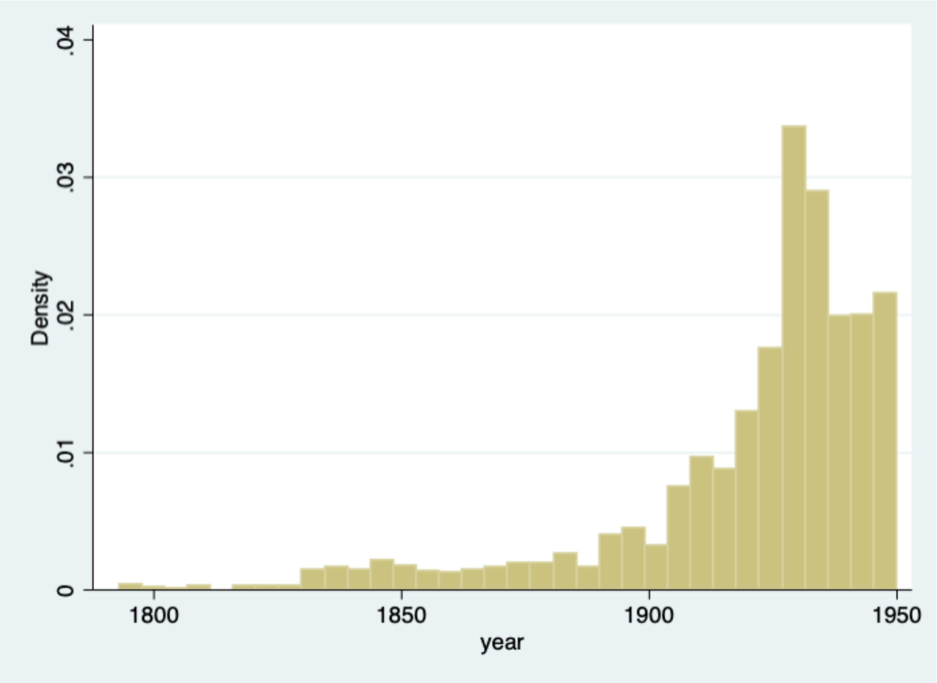
APPENDIX C

CHAPTER3

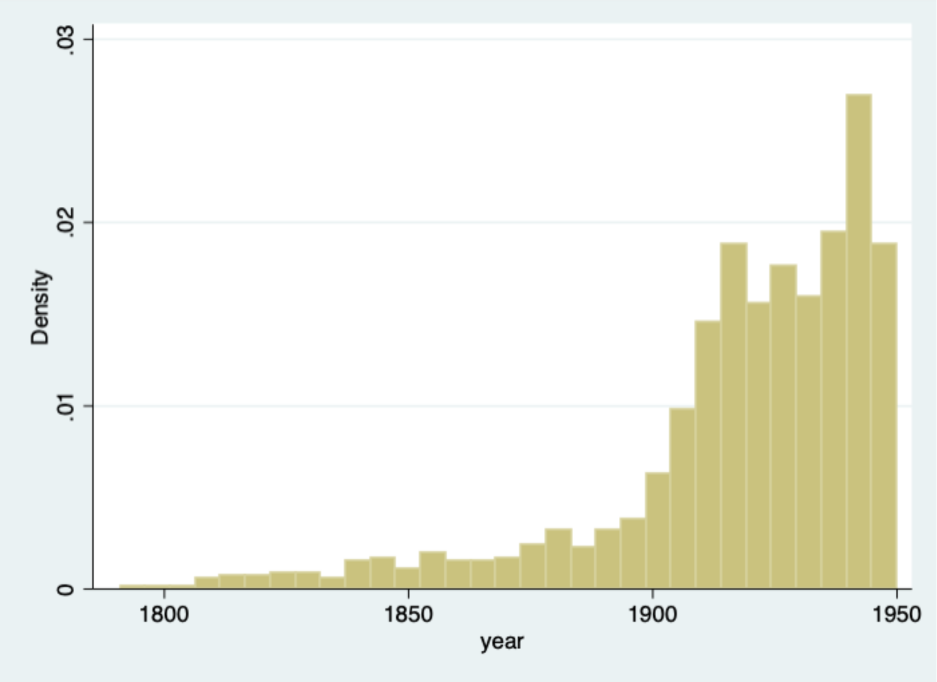
Figure C.1: Computer peripheral operators working on the mainframe IBM360

Figure C.2: Distribution of home economics bibliography by subfield and year

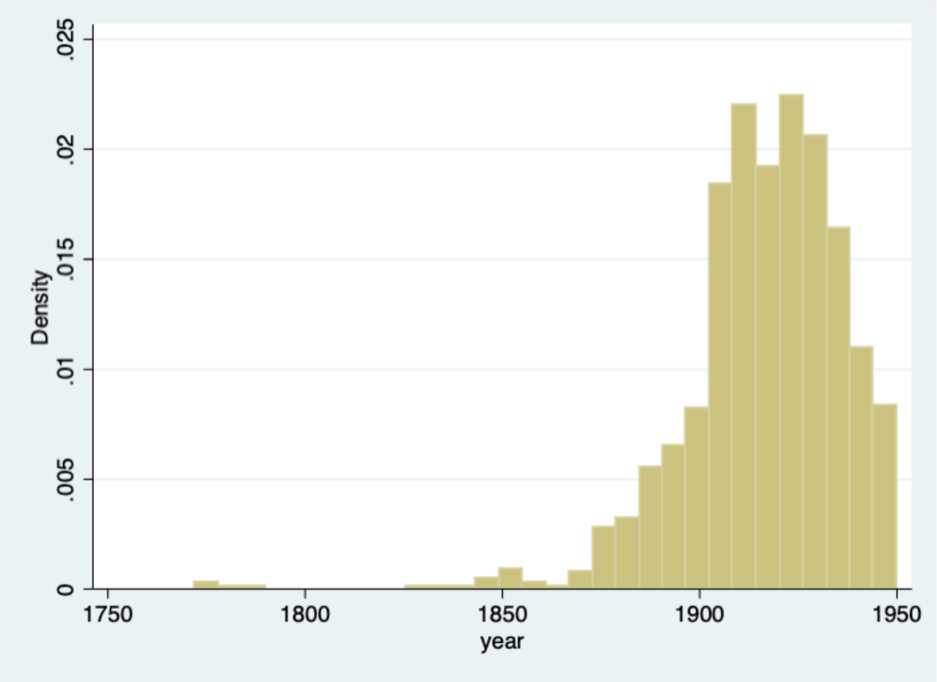
(a) Child care and human development



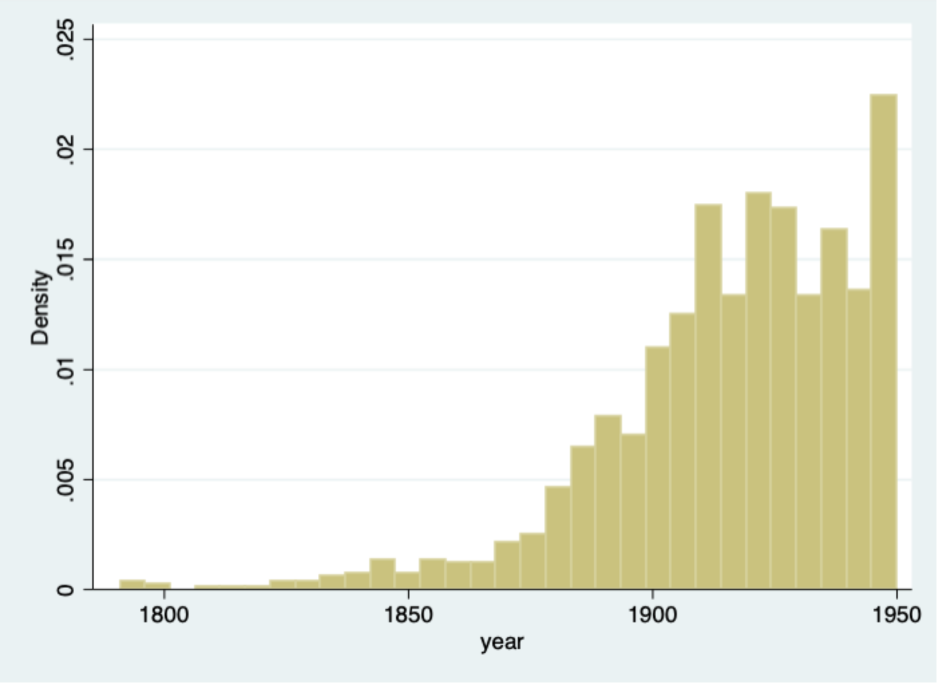
(b) Food & Nutrition



(c) Hygiene

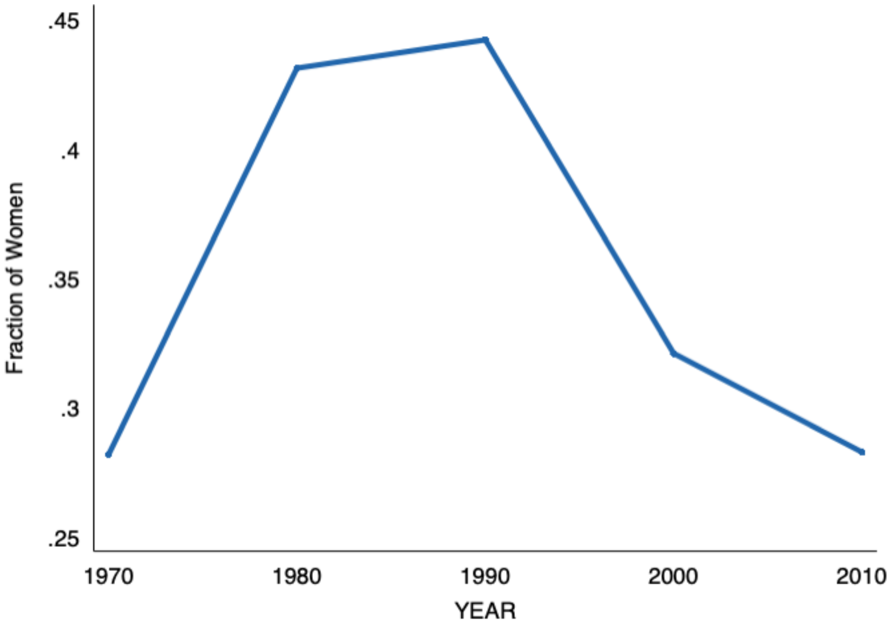


(d) Clothing and textile



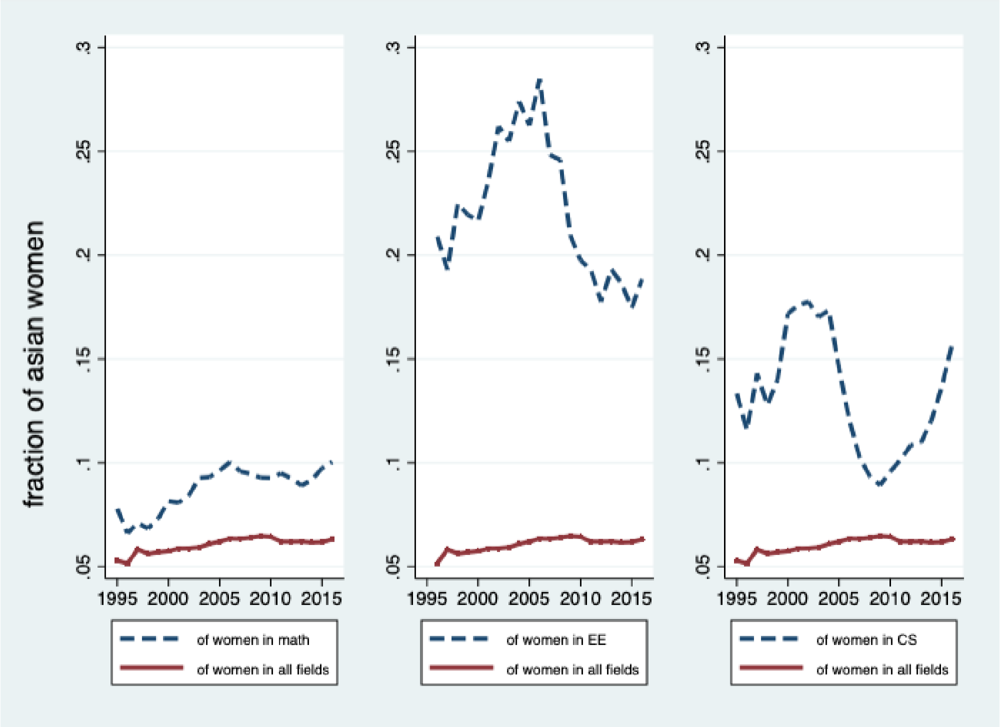
Data Source: (Albert R. Mann Library, 2005).

Figure C.3: Fraction of women in computer industry



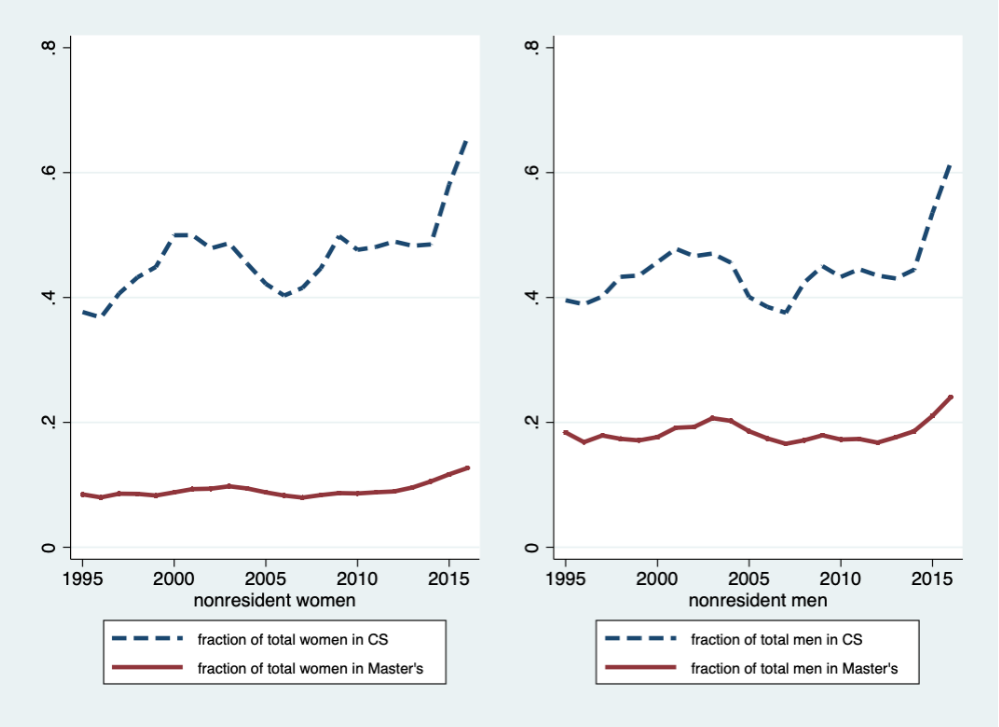
Notes: This figure shows the fraction of women that are employed in the computer industry, working as system analysts & computer scientists, software developers or peripheral operators. *Data sources*: IPUMS random population sample: 1970 1% state fm1, 1980 5% state, 1990 5% state, 2000 5% state, 2010 ACS (Steven Ruggles and Sobek, 2019)

Figure C.4: Asian women's representation in math, EE & CS



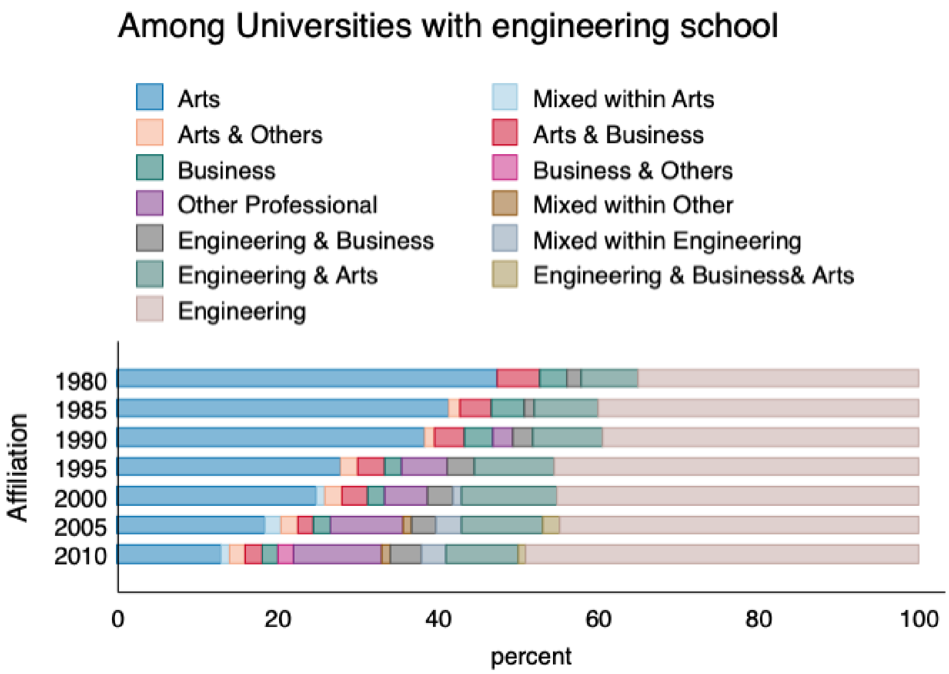
Notes: This figure depicts Asian women's relative representation in math, physical science and engineering. The solid lines plot the fraction of white women among female college graduates across all fields. The dashed lines plot the fraction of white women among female college graduates in math, EE or CS. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

Figure C.5: Foreign students' representation in CS at Master's level



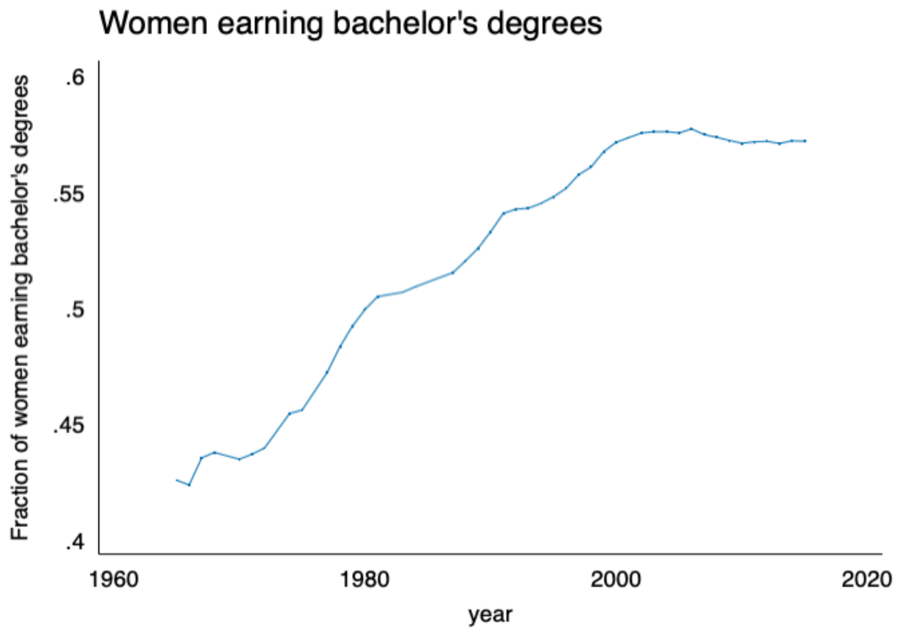
Notes: This figure depicts foreign female and male student's relative representation in CS. The solid lines plot the fraction of nonresident-alien women/men among female/male Master's graduates across all fields. The dashed lines plot the fraction of nonresident-alien women/men among female/male Master's graduates in CS. *Data source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016)

Figure C.6: Computer Science Departmental Affiliation



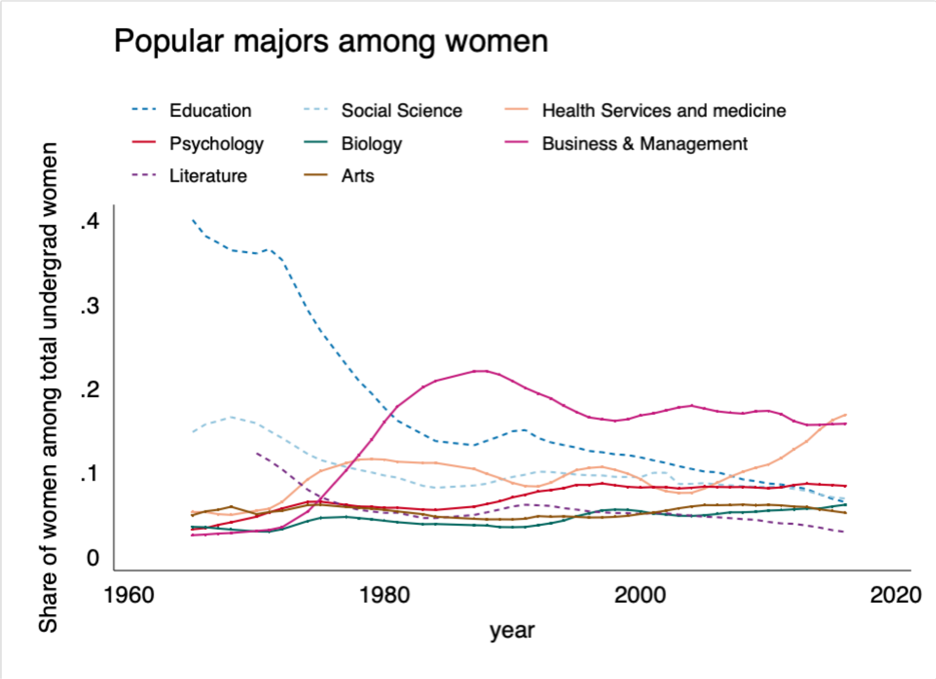
Notes: This figure shows the distribution of CS departmental affiliation from 1980 to 2010, among universities with engineering schools. Affiliation is defined as the schools that offer majors in CS. *Arts* is short for the college of liberal arts & sciences. *Business* is short for the College of Business. *Engineering* is short for the School of Engineering. If a professional school cannot be identified as either engineering or business, it is categorized as *Other Professional* . *Data source*: American College Catalog Study Database (Brint, 2013)

Figure C.7: Trend in the fraction of women earning bachelor's degrees



Notes: This figures plots the fraction of women (women/total graduates) earning bachelor's degrees in the U.S. *Data Source*: IPEDS (Integrated Postsecondary Education Data System, 1986–2016), HEGIS (Higher Education General Information Survey, 1965–1984).

Figure C.8: Rise of popular majors for women since 1970



Notes: This figures plots the share of women (women in a field/total women) earning bachelor's degrees in the U.S.

Figure C.9: Fraction of women earning bachelor's degrees in math, CS and engineering

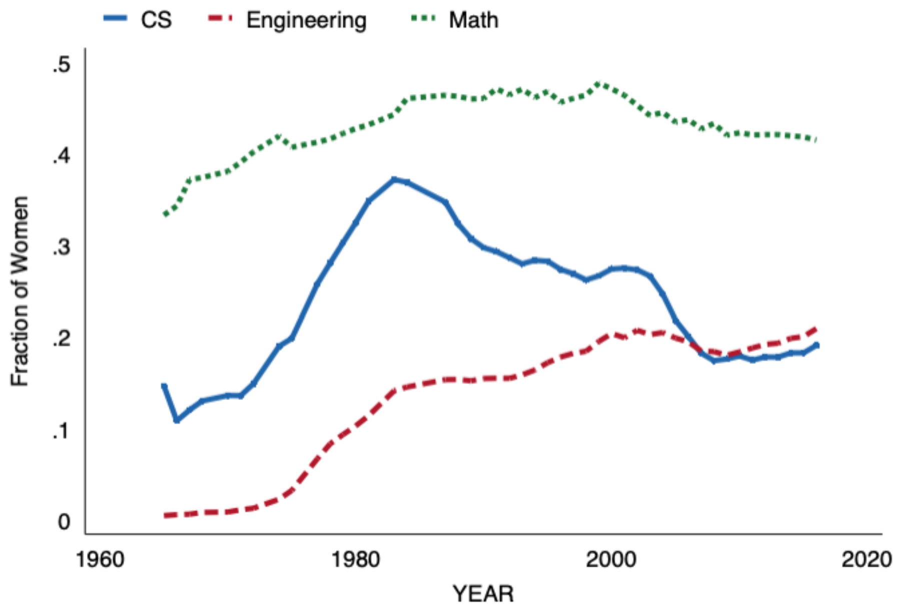


Figure C.10: Fraction of women earning master's degrees in math, CS and engineering

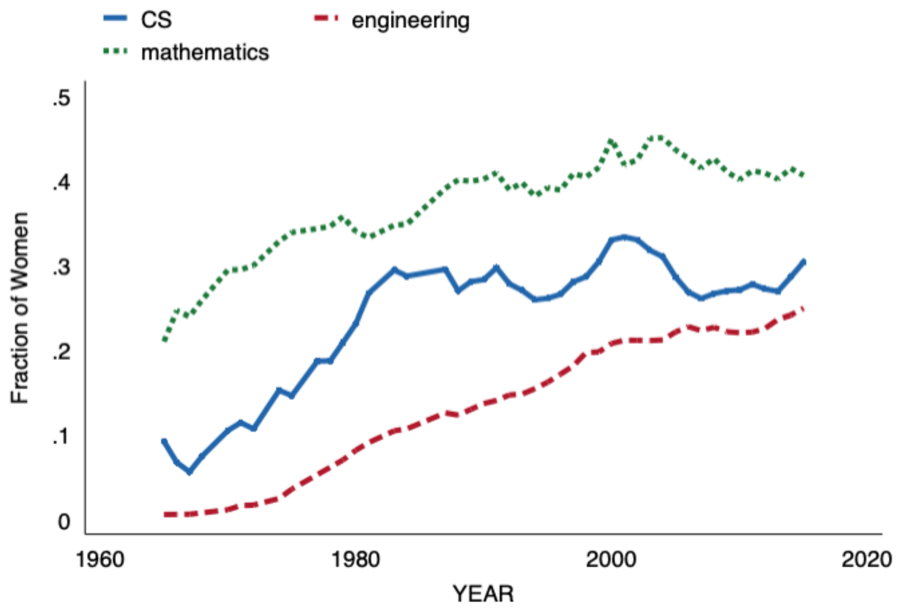


Figure C.11: Share of men and women earning bachelor's degrees in CS

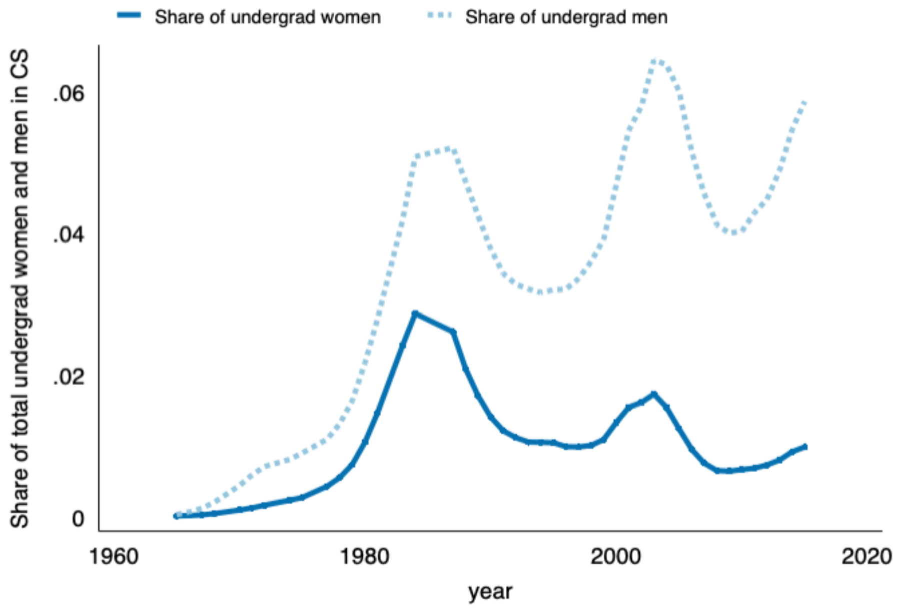


Figure C.12: Number of men and women earning bachelor's degrees in CS

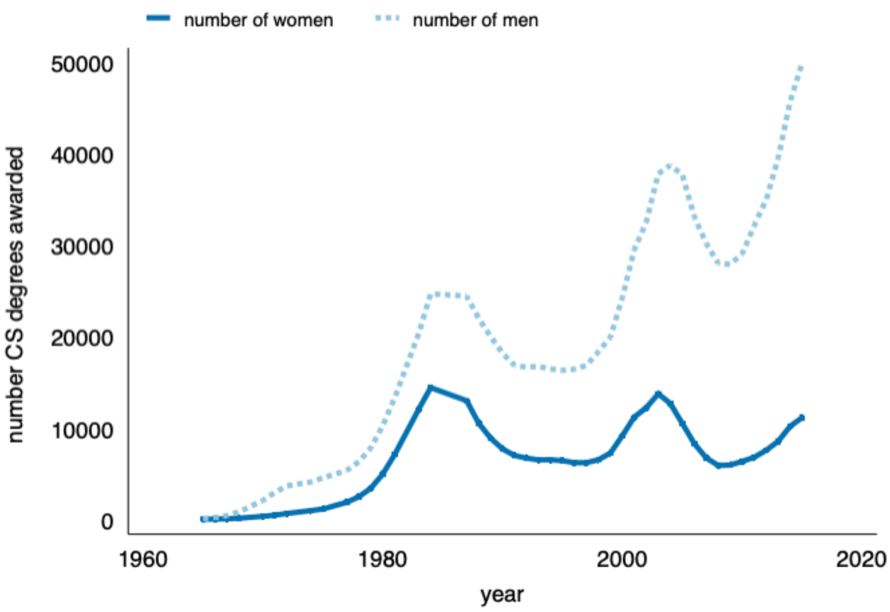
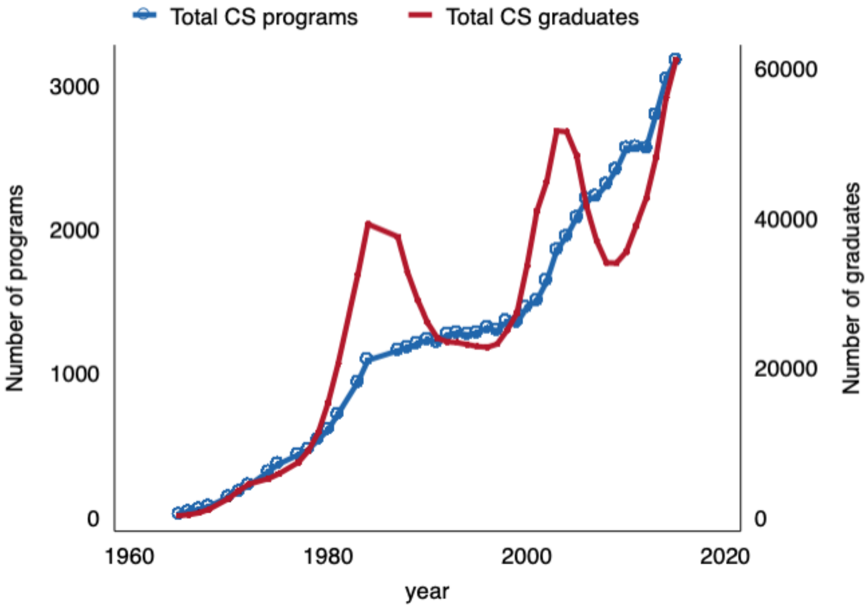
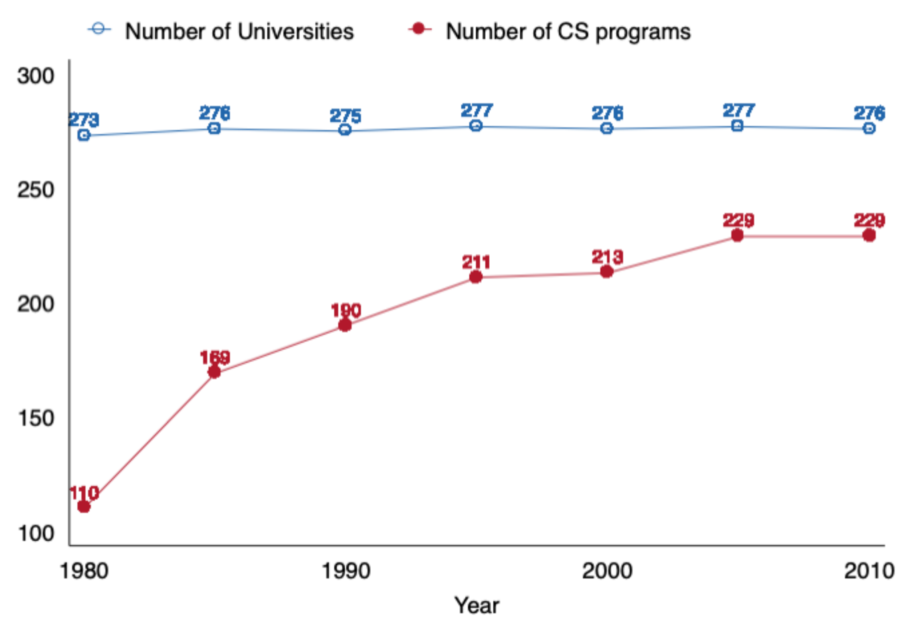


Figure C.13: Trends in total CS programs and CS graduates



Data Source: HEGIS (Higher Education General Information Survey, 1965–1984) and IPEDS (Integrated Postsecondary Education Data System, 1986–2016).

Figure C.14: Total colleges and CS programs in sample



Notes: This figures shows the number of colleges in the core sample and the number of colleges with CS programs. *Data Source*: Brint (2013)

Figure C.15: Correlation between CS and Engineering gender ratio

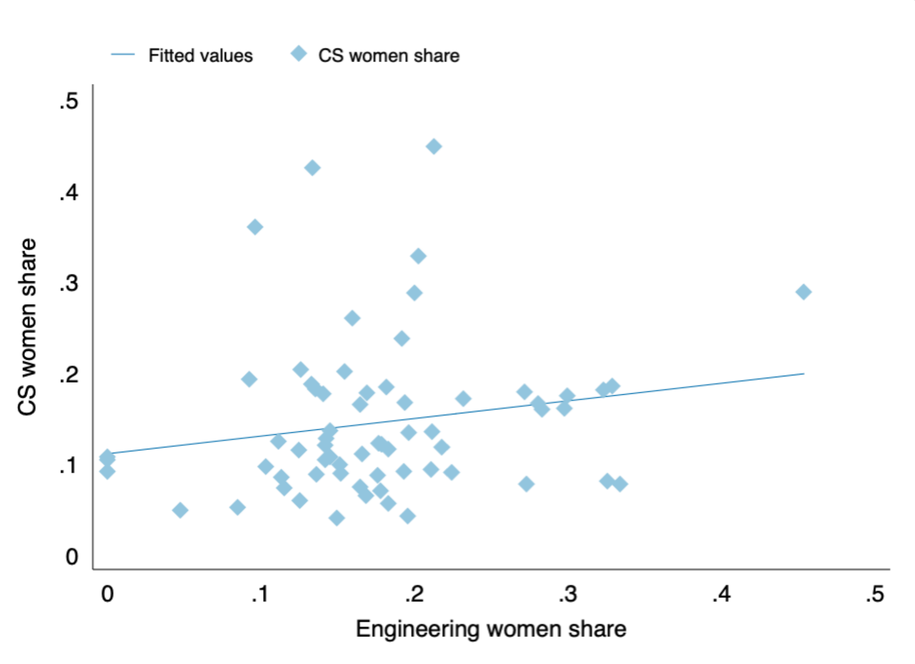
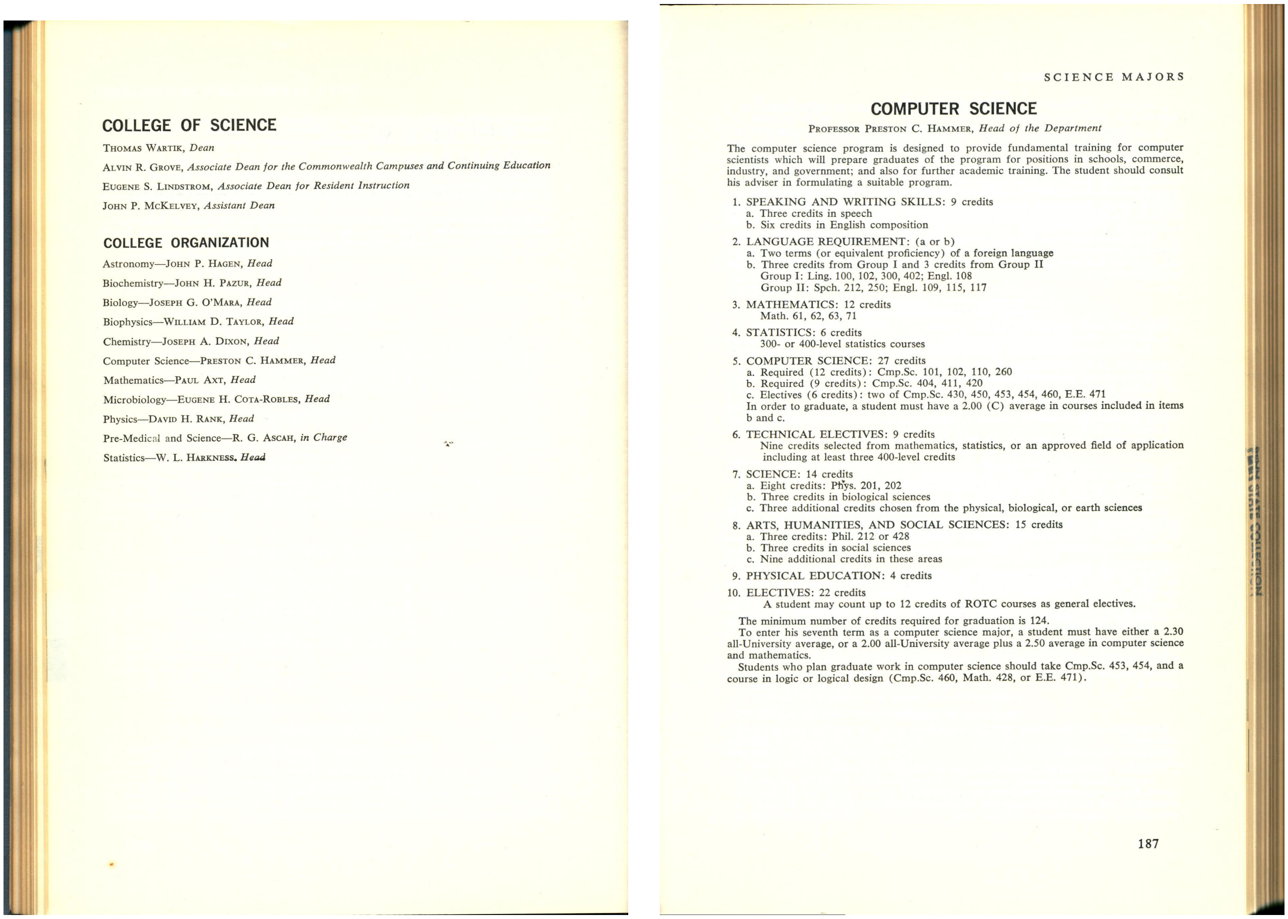


Figure C.16: Sample Course Catalog



COLLEGE OF SCIENCE

THOMAS WARTIK, *Dean*

ALVIN R. GROVE, *Associate Dean for the Commonwealth Campuses and Continuing Education*

EUGENE S. LINDSTROM, *Associate Dean for Resident Instruction*

JOHN P. MCKELVEY, *Assistant Dean*

COLLEGE ORGANIZATION

Astronomy—JOHN P. HAGEN, *Head*

Biochemistry—JOHN H. PAZUR, *Head*

Biology—JOSEPH G. O'MARA, *Head*

Biophysics—WILLIAM D. TAYLOR, *Head*

Chemistry—JOSEPH A. DIXON, *Head*

Computer Science—PRESTON C. HAMMER, *Head*

Mathematics—PAUL AXT, *Head*

Microbiology—EUGENE H. COTA-ROBLES, *Head*

Physics—DAVID H. RANK, *Head*

Pre-Medical and Science—R. G. ASCAH, *in Charge*

Statistics—W. L. HARKNESS. *Head*

SCIENCE MAJORS

COMPUTER SCIENCE

PROFESSOR PRESTON C. HAMMER, *Head of the Department*

The computer science program is designed to provide fundamental training for computer scientists which will prepare graduates of the program for positions in schools, commerce, industry, and government; and also for further academic training. The student should consult his adviser in formulating a suitable program.

1. SPEAKING AND WRITING SKILLS: 9 credits
   a. Three credits in speech
   b. Six credits in English composition
2. LANGUAGE REQUIREMENT: (a or b)
   a. Two terms (or equivalent proficiency) of a foreign language
   b. Three credits from Group I and 3 credits from Group II
      Group I: Ling. 100, 102, 300, 402; Engl. 108
      Group II: Spch. 212, 250; Engl. 109, 115, 117
3. MATHEMATICS: 12 credits
   Math. 61, 62, 63, 71
4. STATISTICS: 6 credits
   300- or 400-level statistics courses
5. COMPUTER SCIENCE: 27 credits
   a. Required (12 credits): Cmp.Sc. 101, 102, 110, 260
   b. Required (9 credits): Cmp.Sc. 404, 411, 420
   c. Electives (6 credits): two of Cmp.Sc. 430, 450, 453, 454, 460, E.E. 471
   In order to graduate, a student must have a 2.00 (C) average in courses included in items b and c.
6. TECHNICAL ELECTIVES: 9 credits
   Nine credits selected from mathematics, statistics, or an approved field of application including at least three 400-level credits
7. SCIENCE: 14 credits
   a. Eight credits: Phys. 201, 202
   b. Three credits in biological sciences
   c. Three additional credits chosen from the physical, biological, or earth sciences
8. ARTS, HUMANITIES, AND SOCIAL SCIENCES: 15 credits
   a. Three credits: Phil. 212 or 428
   b. Three credits in social sciences
   c. Nine additional credits in these areas
9. PHYSICAL EDUCATION: 4 credits
10. ELECTIVES: 22 credits
    A student may count up to 12 credits of ROTC courses as general electives.

The minimum number of credits required for graduation is 124.

To enter his seventh term as a computer science major, a student must have either a 2.30 all-University average, or a 2.00 all-University average plus a 2.50 average in computer science and mathematics.

Students who plan graduate work in computer science should take Cmp.Sc. 453, 454, and a course in logic or logical design (Cmp.Sc. 460, Math. 428, or E.E. 471).

187

Notes: These pages are taken from Pennsylvania State University's 1970 course catalog.

Figure C.17: Degree requirement

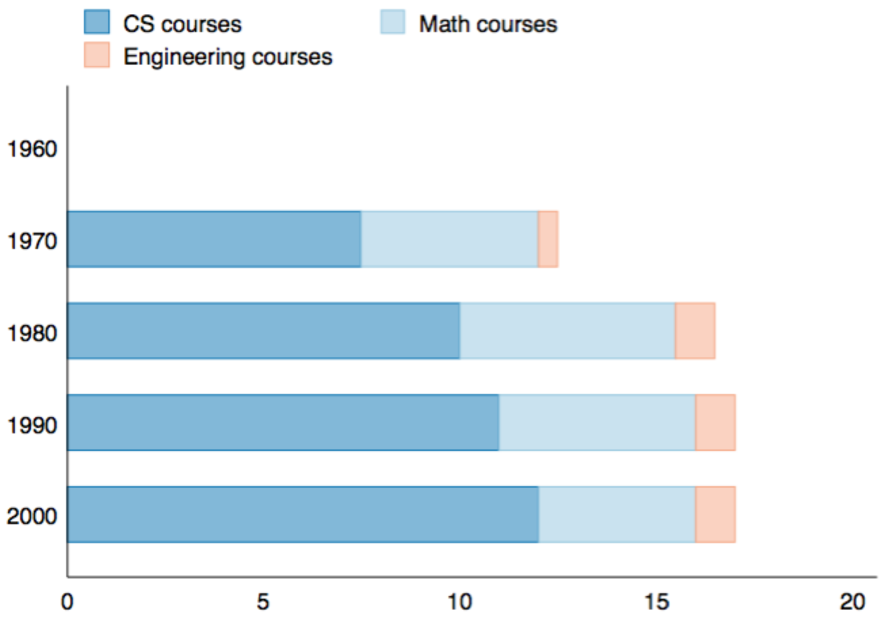


Figure C.18: Fraction of women in various science subjects

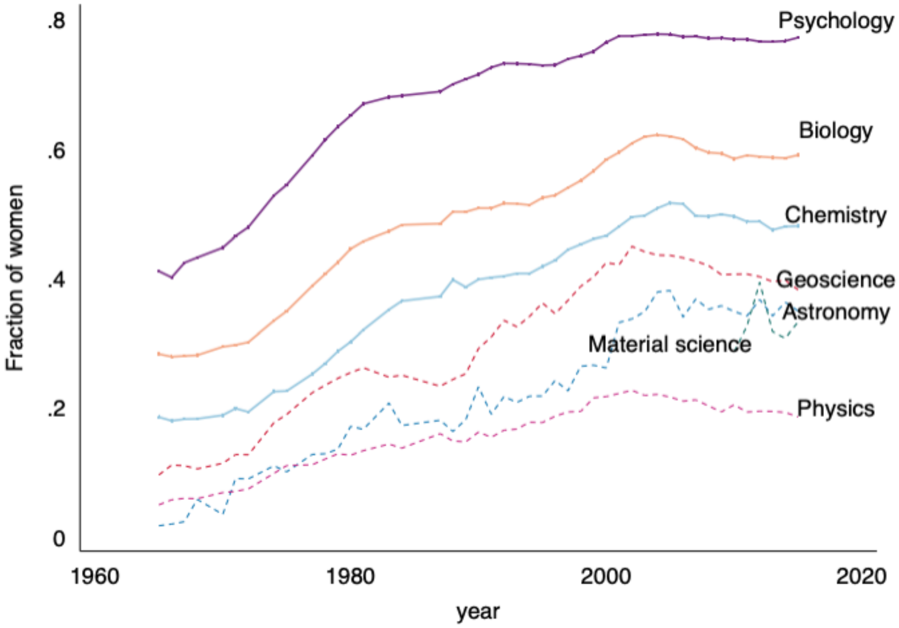


Figure C.19: Fraction of women in various engineering subjects

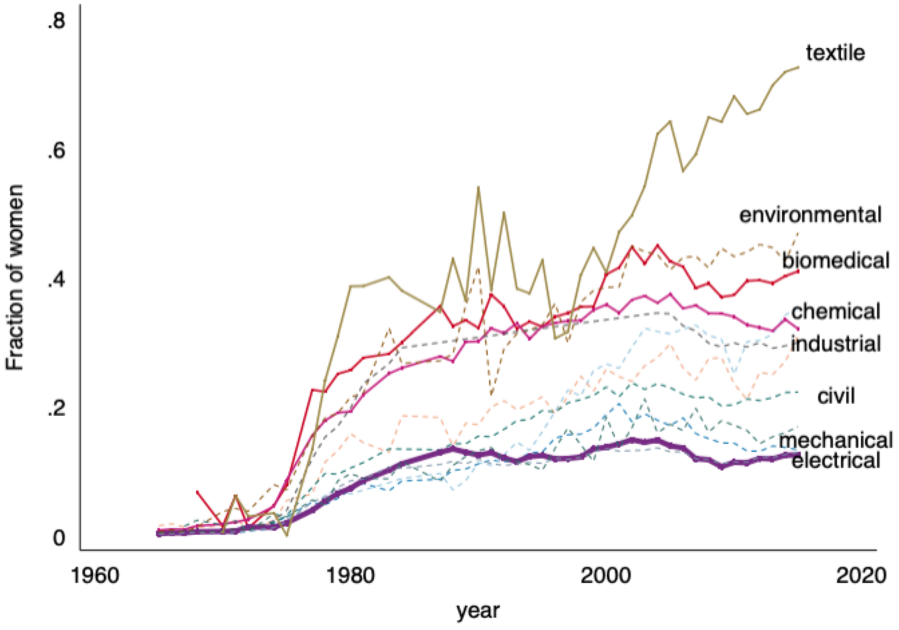
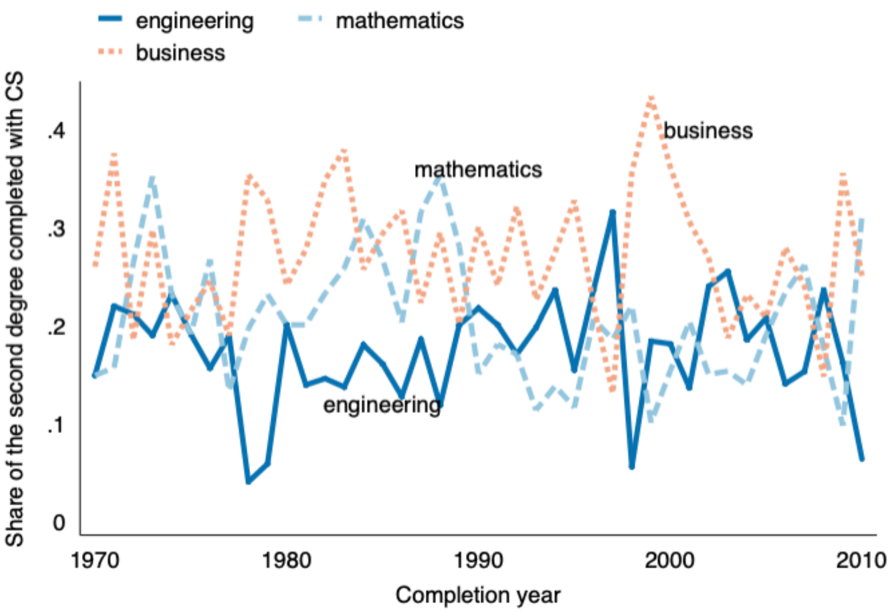
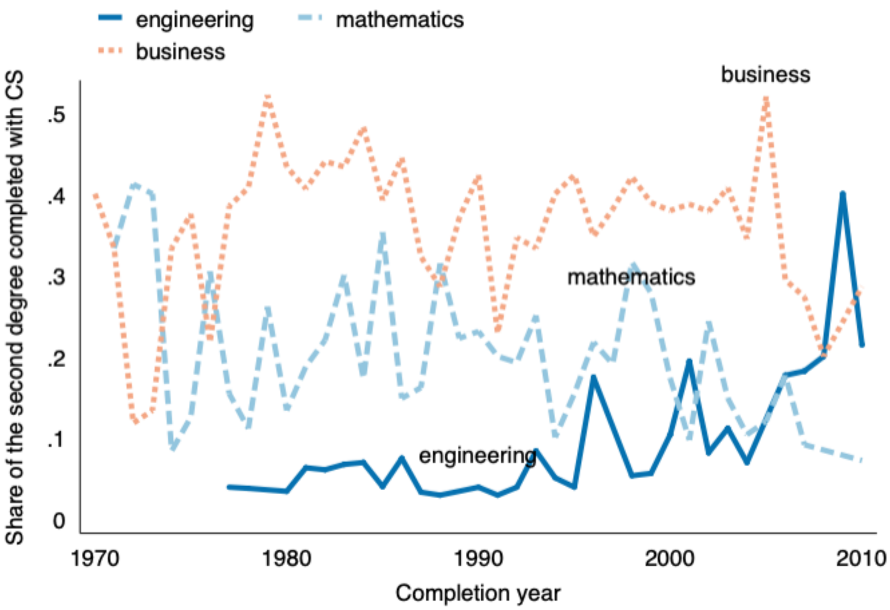


Figure C.20: Double Major Distribution by gender

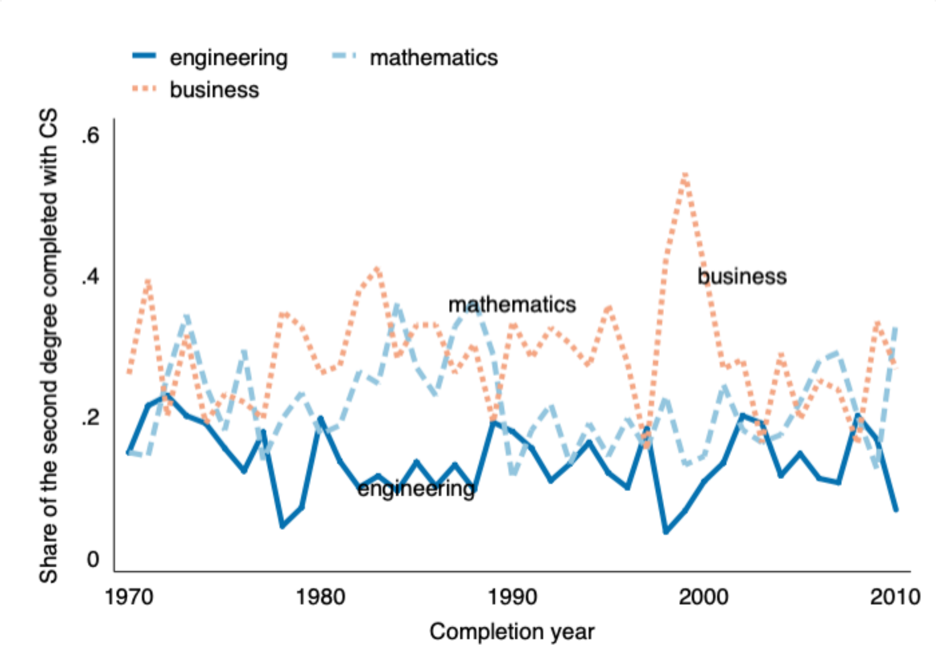
(a) Men



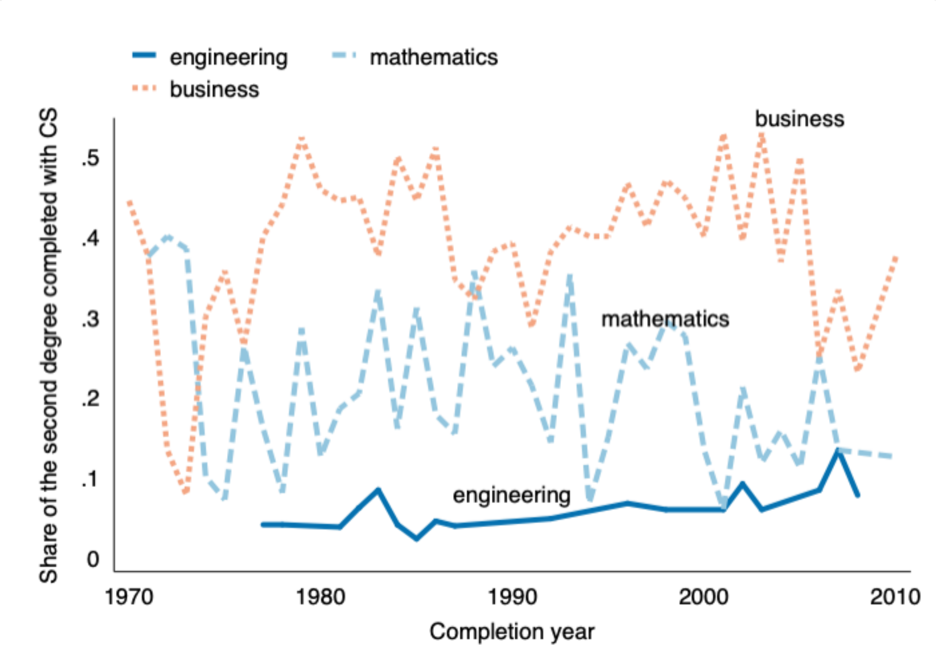
(b) Women



(c) Men, excluding Asian men

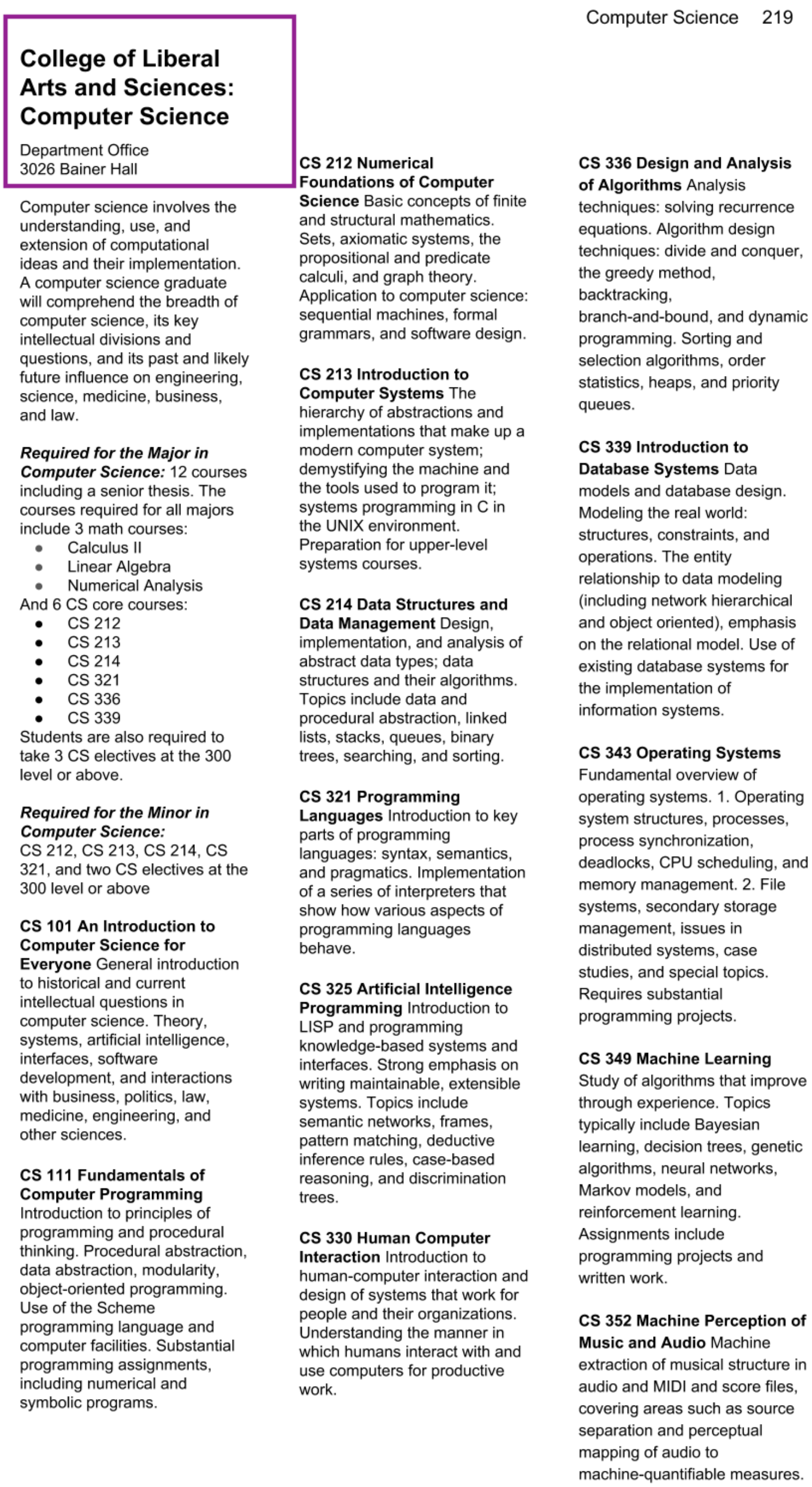


(d) Women, excluding Asian women



Notes: This figure shows the share of the second bachelor's degrees in mathematics, engineering and business completed with CS from 1970 to 2010. The completion year equals 2010 - age + 22, assuming that the age of graduation is 22. *Data Source*: American Community Survey 2010 (Steven Ruggles and Sobek, 2019).

Figure C.21: Syllabus A



Computer Science 219

College of Liberal Arts and Sciences: Computer Science

Department Office
3026 Bainer Hall

Computer science involves the understanding, use, and extension of computational ideas and their implementation. A computer science graduate will comprehend the breadth of computer science, its key intellectual divisions and questions, and its past and likely future influence on engineering, science, medicine, business, and law.

Required for the Major in Computer Science: 12 courses including a senior thesis. The courses required for all majors include 3 math courses:

- Calculus II
- Linear Algebra
- Numerical Analysis

And 6 CS core courses:

- CS 212
- CS 213
- CS 214
- CS 321
- CS 336
- CS 339

Students are also required to take 3 CS electives at the 300 level or above.

Required for the Minor in Computer Science:
CS 212, CS 213, CS 214, CS 321, and two CS electives at the 300 level or above

CS 101 An Introduction to Computer Science for Everyone General introduction to historical and current intellectual questions in computer science. Theory, systems, artificial intelligence, interfaces, software development, and interactions with business, politics, law, medicine, engineering, and other sciences.

CS 111 Fundamentals of Computer Programming Introduction to principles of programming and procedural thinking. Procedural abstraction, data abstraction, modularity, object-oriented programming. Use of the Scheme programming language and computer facilities. Substantial programming assignments, including numerical and symbolic programs.

CS 212 Numerical Foundations of Computer Science Basic concepts of finite and structural mathematics. Sets, axiomatic systems, the propositional and predicate calculi, and graph theory. Application to computer science: sequential machines, formal grammars, and software design.

CS 213 Introduction to Computer Systems The hierarchy of abstractions and implementations that make up a modern computer system; demystifying the machine and the tools used to program it; systems programming in C in the UNIX environment. Preparation for upper-level systems courses.

CS 214 Data Structures and Data Management Design, implementation, and analysis of abstract data types; data structures and their algorithms. Topics include data and procedural abstraction, linked lists, stacks, queues, binary trees, searching, and sorting.

CS 321 Programming Languages Introduction to key parts of programming languages: syntax, semantics, and pragmatics. Implementation of a series of interpreters that show how various aspects of programming languages behave.

CS 325 Artificial Intelligence Programming Introduction to LISP and programming knowledge-based systems and interfaces. Strong emphasis on writing maintainable, extensible systems. Topics include semantic networks, frames, pattern matching, deductive inference rules, case-based reasoning, and discrimination trees.

CS 330 Human Computer Interaction Introduction to human-computer interaction and design of systems that work for people and their organizations. Understanding the manner in which humans interact with and use computers for productive work.

CS 336 Design and Analysis of Algorithms Analysis techniques: solving recurrence equations. Algorithm design techniques: divide and conquer, the greedy method, backtracking, branch-and-bound, and dynamic programming. Sorting and selection algorithms, order statistics, heaps, and priority queues.

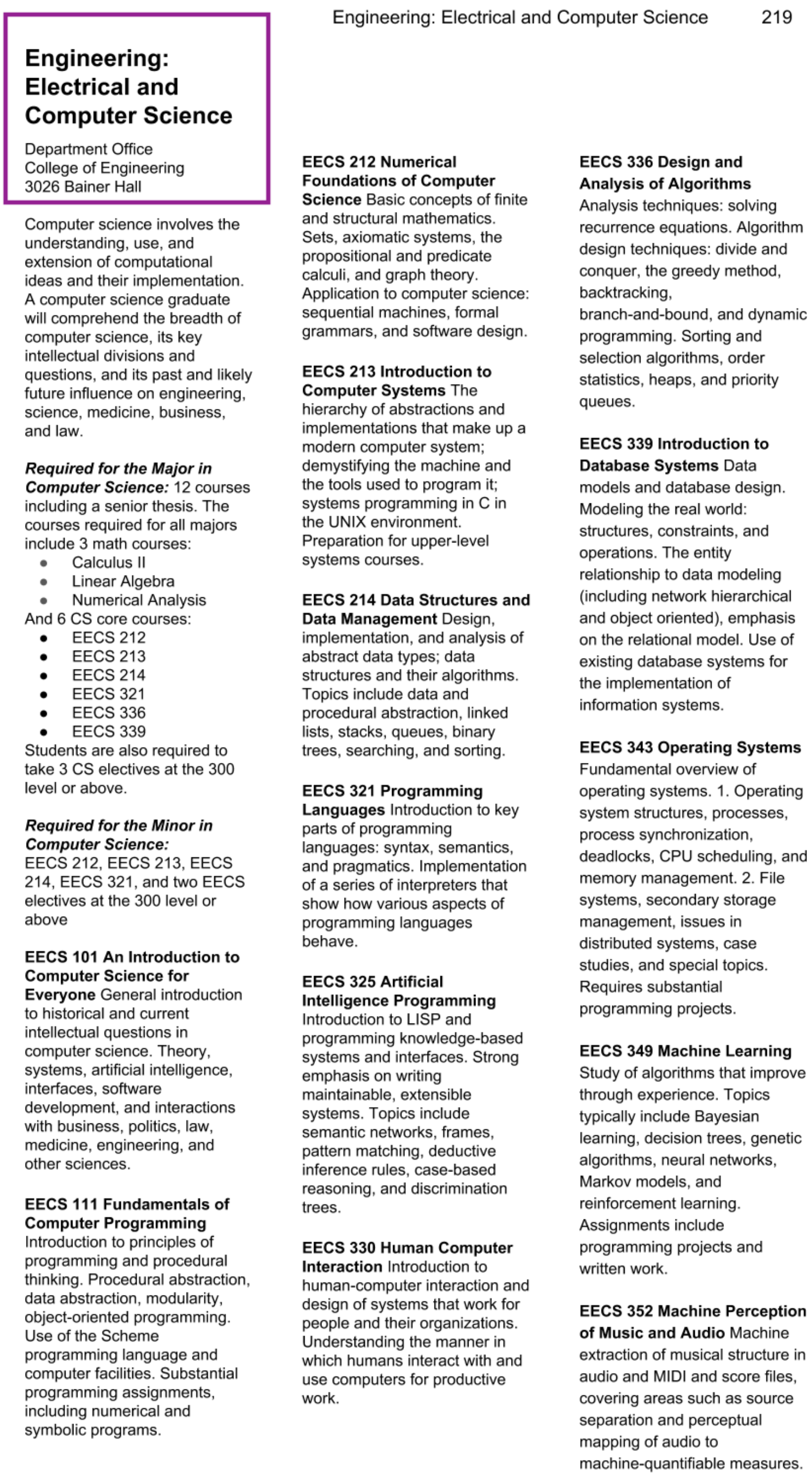
CS 339 Introduction to Database Systems Data models and database design. Modeling the real world: structures, constraints, and operations. The entity relationship to data modeling (including network hierarchical and object oriented), emphasis on the relational model. Use of existing database systems for the implementation of information systems.

CS 343 Operating Systems Fundamental overview of operating systems. 1. Operating system structures, processes, process synchronization, deadlocks, CPU scheduling, and memory management. 2. File systems, secondary storage management, issues in distributed systems, case studies, and special topics. Requires substantial programming projects.

CS 349 Machine Learning Study of algorithms that improve through experience. Topics typically include Bayesian learning, decision trees, genetic algorithms, neural networks, Markov models, and reinforcement learning. Assignments include programming projects and written work.

CS 352 Machine Perception of Music and Audio Machine extraction of musical structure in audio and MIDI and score files, covering areas such as source separation and perceptual mapping of audio to machine-quantifiable measures.

Figure C.22: Syllabus B



Engineering: Electrical and Computer Science 219

Engineering: Electrical and Computer Science

Department Office
College of Engineering
3026 Bainer Hall

Computer science involves the understanding, use, and extension of computational ideas and their implementation. A computer science graduate will comprehend the breadth of computer science, its key intellectual divisions and questions, and its past and likely future influence on engineering, science, medicine, business, and law.

Required for the Major in Computer Science: 12 courses including a senior thesis. The courses required for all majors include 3 math courses:

- Calculus II
- Linear Algebra
- Numerical Analysis

And 6 CS core courses:

- EECS 212
- EECS 213
- EECS 214
- EECS 321
- EECS 336
- EECS 339

Students are also required to take 3 CS electives at the 300 level or above.

Required for the Minor in Computer Science:
EECS 212, EECS 213, EECS 214, EECS 321, and two EECS electives at the 300 level or above

EECS 101 An Introduction to Computer Science for Everyone General introduction to historical and current intellectual questions in computer science. Theory, systems, artificial intelligence, interfaces, software development, and interactions with business, politics, law, medicine, engineering, and other sciences.

EECS 111 Fundamentals of Computer Programming Introduction to principles of programming and procedural thinking. Procedural abstraction, data abstraction, modularity, object-oriented programming. Use of the Scheme programming language and computer facilities. Substantial programming assignments, including numerical and symbolic programs.

EECS 212 Numerical Foundations of Computer Science Basic concepts of finite and structural mathematics. Sets, axiomatic systems, the propositional and predicate calculi, and graph theory. Application to computer science: sequential machines, formal grammars, and software design.

EECS 213 Introduction to Computer Systems The hierarchy of abstractions and implementations that make up a modern computer system; demystifying the machine and the tools used to program it; systems programming in C in the UNIX environment. Preparation for upper-level systems courses.

EECS 214 Data Structures and Data Management Design, implementation, and analysis of abstract data types; data structures and their algorithms. Topics include data and procedural abstraction, linked lists, stacks, queues, binary trees, searching, and sorting.

EECS 321 Programming Languages Introduction to key parts of programming languages: syntax, semantics, and pragmatics. Implementation of a series of interpreters that show how various aspects of programming languages behave.

EECS 325 Artificial Intelligence Programming Introduction to LISP and programming knowledge-based systems and interfaces. Strong emphasis on writing maintainable, extensible systems. Topics include semantic networks, frames, pattern matching, deductive inference rules, case-based reasoning, and discrimination trees.

EECS 330 Human Computer Interaction Introduction to human-computer interaction and design of systems that work for people and their organizations. Understanding the manner in which humans interact with and use computers for productive work.

EECS 336 Design and Analysis of Algorithms Analysis techniques: solving recurrence equations. Algorithm design techniques: divide and conquer, the greedy method, backtracking, branch-and-bound, and dynamic programming. Sorting and selection algorithms, order statistics, heaps, and priority queues.

EECS 339 Introduction to Database Systems Data models and database design. Modeling the real world: structures, constraints, and operations. The entity relationship to data modeling (including network hierarchical and object oriented), emphasis on the relational model. Use of existing database systems for the implementation of information systems.

EECS 343 Operating Systems Fundamental overview of operating systems. 1. Operating system structures, processes, process synchronization, deadlocks, CPU scheduling, and memory management. 2. File systems, secondary storage management, issues in distributed systems, case studies, and special topics. Requires substantial programming projects.

EECS 349 Machine Learning Study of algorithms that improve through experience. Topics typically include Bayesian learning, decision trees, genetic algorithms, neural networks, Markov models, and reinforcement learning. Assignments include programming projects and written work.

EECS 352 Machine Perception of Music and Audio Machine extraction of musical structure in audio and MIDI and score files, covering areas such as source separation and perceptual mapping of audio to machine-quantifiable measures.

Table C.1: Major choice in Computer Science

Stated choice	(1)	(2)	(3)	(4)	(5)
Completion in 4 years	0.5080*				
	(0.2613)				
GPA $\geq$ 3.0		-0.0434			
		(0.1775)			
Enjoyment of coursework			0.0783**		
			(0.0365)		
Family approval				0.2924	
				(0.3126)	
Job upon graduation					0.2471
					(0.2415)
Observations	29	29	29	29	29
Log likelihood	-14.474738	-17.051623	-14.583911	-16.577907	-16.51501

Table C.2: Effect of engineering syllabus on expected enjoyment from coursework

Expectation on the enjoyment of coursework	(1)	(2)	(3)	(4)	(5)	(6)
Engineering labeling	-1.3056*	-1.2285*	-1.2610*	-1.3121*	-1.2569*	-1.3155*
	(0.6462)	(0.6395)	(0.6330)	(0.6761)	(0.6324)	(0.6632)
Math interest		0.8162				
		(0.6097)				
GPA			0.4909			
			(0.3313)			
Grade				-0.0093		
				(0.2177)		
Expected % women in CS					-0.0248	
					(0.0165)	
Preferred gender composition						0.0180
						(0.1467)
Observations	29	29	29	29	29	29
R^2	.131338	.187352	.198998	.131399	.200879	.131841