Essays on Science and Innovation

by

Jungkyu Suh

Department of Business Administration
Duke University

Date:_____

Approved:

_____
Ashish Arora, Co-Supervisor

_____
Sharon Belenzon, Co-Supervisor

_____
Wesley Cohen

_____
Daniel Gross

_____
David Robinson

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of Business Administration
in the Graduate School of Duke University

2022

ABSTRACT

Essays on Science and Innovation

by

Jungkyu Suh

Department of Business Administration
Duke University

Date:_____

Approved:

_____
Ashish Arora, Co-Supervisor

_____
Sharon Belenzon, Co-Supervisor

_____
Wesley Cohen

_____
Daniel Gross

_____
David Robinson

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of Business Administration
in the Graduate School of Duke University

2022

Copyright © 2022 by Jungkyu Suh
All rights reserved.

Abstract

The commercialization of scientific discoveries into innovation has traditionally been the purview of large corporations operating central R&D laboratories through much of the past century. The past four decades have seen this model being gradually supplanted by a more decentralized system of universities and VC-backed startups that have displaced large corporations as the conductors of scientific research. This dissertation tries to understand how firms create and exploit scientific knowledge in this changing structure of American innovation. The first study examines how scientific knowledge can expand markets for technology and thereby encourage the entry of new science-based firms into invention. The argument is tested in the context of the U.S. patent market and finds that patents citing scientific articles tend to be traded more often, even after controlling for various proxies of patent quality. The second study explores why some American firms started investing in scientific research in the early twentieth century. The chapter relies on a newly assembled panel dataset of innovating firms consisting of their investments in science, patenting, financials and ownership between 1926 and 1940. The empirical patterns reveal that the beginnings of corporate research in America were driven by companies at the technological frontier attempting to take advantage of opportunities for innovation made possible by scientific advances. This investment was especially pronounced for firms based in scientific fields that were underdeveloped in the United States. The final study asks why startups are more likely to bring scientific advances to

market. The existing literature has explained the higher innovative propensity of some startups by their superior scientific capabilities. However, it is also possible that the apparent innovativeness of startups may be a result of firm choice, rather than inherent capability gaps with respect to incumbents. Startups may choose novel products that are riskier but offer higher payoffs because they pay a higher entry cost in the form of investments in new factories, sales and distribution channels. I test this entry cost mechanism in the context of the American laser industry which responded to an exogenous influx of Soviet laser science following the end of the Cold War.

to parents, with love

Contents

Abstract **iv**

Dedication **vi**

List of Tables **xii**

List of Figures **xvi**

Acknowledgements **xviii**

1 Introduction **1**

1.1 The Changing Structure of American Innovation 4

1.1.1 Decline of Corporate Science 4

1.1.2 The Rise of Academic and Startup Inventions 6

1.1.3 The Growth of Markets for Technology 7

1.2 Outline of the Dissertation 10

2 Science and the Market for Technology **13**

2.1 Introduction 13

2.2 Science and the Precision of Inventive Claims 19

2.3 A Model of Trade in Patents 23

2.3.1 Probability of Trade 24

2.3.2 Role of Science 25

2.3.3 Novelty of Invention 26

2.3.4 Equilibrium: Specialization and Division of Innovative Labor . 28

2.4 Data . . . 31
2.5 Econometric Framework . . . 35
2.5.1 Baseline Trade Equation (OLS) . . . 35
2.5.2 Instrumental Variable Strategy and Other Robustness Analyses 37
2.6 Estimation Results . . . 42
2.6.1 Reliance on Science and Trade . . . 42
2.6.2 Baseline Trade Equation . . . 44
2.6.3 Entry and Market Structure (OLS) . . . 57
2.6.4 Instrumental Variable Estimation . . . 59
2.7 Discussion and Conclusion . . . 61
3 The Rise of Scientific Research in Corporate America 63
3.1 Introduction . . . 63
3.2 The Rise of Corporate Science in America . . . 70
3.2.1 Causes Behind Corporate Science . . . 72
3.3 Theoretical Framework . . . 75
3.3.1 Model Predictions and Empirical Implications . . . 76
3.4 Data . . . 78
3.4.1 Scientific Publications . . . 79
3.4.2 Corporate Labs . . . 82
3.4.3 Gap in University Science . . . 83
3.4.4 Patents . . . 87
3.4.5 Corporate Ownership, Financial Statements and Industry Concentration . . . 88
3.4.6 Descriptive Statistics . . . 90
3.5 Empirical Results . . . 93
3.5.1 Who Invests in Science? . . . 93

3.5.2 Interaction with Gaps in University Science (OLS) 95
3.5.3 Performance Consequences of Corporate Science 100
3.6 Discussion and Conclusion 106
4 Science, Startups and Novelty **109**
4.1 Introduction 109
4.2 Theoretical Framework: Science and Novelty of Innovation 114
4.2.1 A simple model of science commercialization 119
4.3 Empirical Context: Laser Science Across the Iron Curtain 121
4.3.1 Reallocation of Cold War Defense Spending 126
4.4 Data 127
4.4.1 Science 128
4.4.2 Invention 130
4.4.3 Entry 134
4.4.4 Products 135
4.4.5 Exit 138
4.5 Results 139
4.5.1 Stylized Facts on Soviet Science, Inventions and Innovations . 141
4.5.2 Invention (Patents) 146
4.5.3 Entry 149
4.5.4 Innovation (Products) 151
4.5.5 Exit 156
4.6 Discussion and Conclusion 159
5 Conclusions **163**
A Appendix to Chapter 2 **166**
A.1 Data 167

A.1.1 Sample Construction 167
A.1.2 Instrumental Variable Construction 173
A.1.3 IVE Robustness Check 181
A.2 Auxiliary Results 185
B Appendix to Chapter 3 **187**
B.1 Model 188
B.1.1 Setup 188
B.1.2 Stage 2: Innovation 190
B.1.3 Stage 1: Research 191
B.1.4 Result 1: Innovation Leadership 192
B.1.5 Public Science 192
B.1.6 Internal Research and Public Science 194
B.2 Auxiliary Results 197
B.3 Details on Data Construction 203
B.3.1 Corporate Historical Documents and Data Sources 203
B.3.2 Corporate Histories 206
B.3.3 Control Chains 206
B.3.4 Ultimate Controlling Shareholders 207
B.3.5 Matching Corporations to Patents, Publications and Labs . . 207
B.4 Details on Scientific Gap Calculations 213
B.4.1 American Men of Science Directory 213
B.4.2 Web of Science Affiliations Coding 215
B.4.3 Journal Country Coding and Citation Flows 217
B.4.4 Comparison Between Gap Measures 226

C Appendix to Chapter 4 **230**

C.1 A Model of Technology Commercialization 231

C.1.1 Payoffs and Probabilities of Experimentation, Entry, and Exit 231

C.1.2 Key findings 233

C.2 Historical Details on Laser Science in the Soviet Union 236

C.3 Data Appendix 241

C.3.1 Details on Product and Market Data from *Laser Focus* 241

C.3.2 Federal Laser Contracts 244

C.3.3 Soviet Scientist Immigration to the United States 247

C.4 Robustness Checks 251

C.4.1 Auxiliary Patent-Level Results 251

C.4.2 Replication of Results without SDI-related Lasers 252

C.4.3 Interaction with Soviet S&E Immigration to American Cities . 256

Bibliography **262**

Biography **282**

List of Tables

1.1 U.S. Distribution of technology licensing receipts by sector for 2002 & 2011 (in billions of dollars) 9

2.1 Summary Statistics for Main Variables 35

2.2 Mean Comparisons of Patent Characteristics, by Science and Reassignment Status 43

2.3 Markets for Technology and Reliance on Science (OLS) 46

2.4 Markets for Technology and Reliance on Science, Alternative Measures (OLS) 47

2.5 Reliance on Science, for Inductees of the National Inventors Hall of Fame 49

2.6 Science and MFT, by Seller Size and Buyer Heterogeneity 51

2.7 Science and MFT, by Patent Novelty (OLS) 54

2.8 Textual Similarity to Science and MFT, by Seller Size, Buyer Heterogeneity, and Patent Novelty (OLS) 56

2.9 Science, MFT, and entry into invention 58

2.10 Post Cold War Federal R&D Shifts and MFT (IV Estimates) 60

3.1 Corporate Scientific Publications, by OECD Subfield 81

3.2 Summary Statistics of Main Variables 90

3.3 Investment in Corporate Science, by Firm Characteristics 95

3.4 Corporate Publications and Gaps in University Science, by Proximity to Technological Frontier 97

3.5 Corporate Publications and Gaps in University Science, by Size and Corporate Structure 99

3.6 Corporate Science and "Home-Run" Patents 103

3.7 Corporate Science and Market-to-Book Ratios 105

4.1 Empirical Predictions 120

4.2 Summary of Variables 128

4.3 Soviet Paper Share by Gain Medium (1980-90) 130

4.4 Descriptive Statistics for Laser Patents 132

4.5 Descriptive Statistics (Firm-Year) 135

4.6 Novelty of Laser Products 144

4.7 Novelty of Laser Patents (CPC: H01S) and Citations to Soviet Science, by Assignee Type 146

4.8 Citations to Soviet Laser Journals and Gain Media's Soviet Paper Share (H01S Patents) 148

4.9 Entry of Laser Patentees into Product Market 151

4.10 Startup Entry into Novel Products 154

4.11 Startup Entry into Novel Products, by Downstream Conditions . . . 156

4.12 Survival Rates for Producers of "Soviet" Lasers 158

A.1 Technological Sector Classifications, by 2 Digit IPCs) 173

A.2 Government Funding of Procurement Contracts for R&D Services . . 175

A.3 Post Cold War Federal R&D Shifts and MFT (Predicted Funding Shocks) 182

A.4 Reliance on Science and MFT, Forward Patent Citation Controls) . . 185

A.5 Post Cold War Federal R&D Shifts and MFT (Bootstrapped Standard Errors) 186

B.1 Corporate Publications and Gaps in University Science (Scientists), by Firm Characteristics 199

B.2 Corporate Publications and Gaps in University Science (Cites), by Firm Characteristics 200

B.3 Corporate Science and "Home-Run" Patents (AMS Gaps) 201

B.4 Corporate Science and "Home-Run" Patents (WOS Citation Gaps) . 202

B.5 Corporate Science and Market-to-Book Ratios 203

B.6 American Corporate Publications (Top 20) 210

B.7 American Corporate Publications, by Scientific Field 211

B.8 WoS Countries and Regions (1/2) 216

B.9 WoS Countries and Regions (2/2) 217

B.10 Country Coding of Journals (1/7) 219

B.11 Country Coding of Journals (2/7) 220

B.12 Country Coding of Journals (3/7) 221

B.13 Country Coding of Journals (4/7) 222

B.14 Country Coding of Journals (5/7) 223

B.15 Country Coding of Journals (6/7) 224

B.16 Country Coding of Journals (7/7) 225

B.17 Number of Scientists and Papers, Europe vs America 228

C.1 Laser Product Types (Construction and Beam Type by Gain Medium) (1/2) 242

C.2 Laser Product Types (Construction and Beam Type by Gain Medium) (2/2) 243

C.3 Top and Bottom 5 MSAs, by Increase in Soviet S&E Migrants 248

C.4 Novelty of Laser Patents (CPC: H01S) and Gain Media, by Assignee Type 251

C.5 Novelty of Laser Patents (CPC: H01S), by Citations to Soviet Science 252

C.6 Novelty of Laser Products (excluding SDI lasers) 254

C.7 Startup Entry into Novel Products (excluding SDI lasers) 255

C.8 Novelty of Laser Patents (CPC: H01S), by Citations to Soviet Science (excluding SDI lasers) 256

C.9 Soviet Scientist Immigration and Lasers (City Level) 257

C.10 Patent Citations to Soviet Laser Science 258

C.11 Novelty of Laser Products, by Soviet S&E Immigration 260

C.12 Startup Entry into Novel Products, by Soviet S&E Immigration . . . 261

List of Figures

1.1 U.S. Scientific Investment and Output (1980-2013) 2

1.2 Business funded and performed research in the United States (1953-2015) 3

1.3 Patent Applications and Gross Licensing Income by Universities (1995-2015) 6

2.1 Timing Structure of Instrumental Variable Estimation 38

2.2 Government Contract R&D Funding, During and After the Cold War 40

3.1 Number of Natural Science Nobel Prize Laureates, by Citizenship at Award 73

3.2 Firm Share of Publications and Personnel 82

3.3 Emergence of Corporate Science, 1926-1940 91

3.4 Heterogeneity of Corporate Science 92

3.5 Corporate Science vs Gaps in University Science, by Industry 93

4.1 Annual Product Type Introductions, by Soviet Paper Share 140

4.2 Product Novelty Before & After 1990, by Soviet Paper Share 141

4.3 Entry of Patentees into Product Market 149

4.4 Product Novelty and Firm Type, Before and After Soviet Collapse . . 152

4.5 New Downstream Applications by Firm Type (Post-1990, Soviet-type Products) 155

A.1 Markets for Technology and Production and Use of Science 167

A.2 U.S. R&D Funding, by Source 179

A.3 Funding Shocks for R&D vs non-R&D Procurements 180
A.4 Share of Patents Traded vs Shock 183
B.1 Corporate Science vs Gaps in University Science, by Industry 197
B.2 A Moody's Manuals entry: The Porto Rico Telephone Company, 1949 204
B.3 Number of Published Utility Patents, 1925-1950 208
B.4 1933 IRL Entry for American Beet & Sugar Company 212
B.5 1933 IRL Entry for AT&T Bell Labs 213
B.6 American Men of Science Entry for Gilbert Lewis (1921) 215
B.7 Comparison of Gaps in University Science (1/2) 229
B.8 Comparison of Gaps in University Science (2/2) 229
C.1 Department of Defense Procurement Contract Obligations in Laser Products (PSC: 5860) 245
C.2 INS Occupation Coding for "Immigrants Admitted to the United States" (1983-1997) 249

Acknowledgements

I thank my supervisors Ashish Arora and Sharon Belenzon for pushing me to be better. My hope is that this dissertation marks the beginning of a scholarship worthy of their patient investment. I thank my committee members Wesley Cohen, Daniel Gross, and David Robinson for expert advice, and my coauthors Andrea Patacconi, Konstantin Kosenko, and Yishay Yafeh for broadening my horizons. I also thank fellow doctoral students in the Strategy Department for support and friendship. Finally, I thank Claire and Elizabeth for giving light, warmth and purpose to my doctorate. Research can feel like pushing a boulder up a hill. I was fortunate not to have pushed alone.

1

Introduction

> "...without scientific progress no amount of achievement in other directions can insure our health, prosperity, and security as a nation in the modern world." (Bush, 1945)

A defining feature of modern economic growth is the systematic application of science to advance technology.[1] Many innovations that spurred economic growth in the twentieth century, including synthetic fibers, plastics, integrated circuits, and gene therapy, originated from advances in the natural sciences, engineering and medicine. Science, by producing "a potential for technology far greater than existed previously," clearly distinguishes modern economic growth from previous economic epochs (Kuznets, 1971).

However, despite sustained increases in the quantity of scientific knowledge, productivity growth in most advanced economies has stagnated in recent decades in comparison to a "golden age" in the mid-twentieth century. Using data from the United States, Gordon (2016) shows that real GDP per hour (i.e., labor produc-

[1] *This chapter adapts material from joint work with Ashish Arora, Sharon Belenzon, and Andrea Patacconi (Arora et al., 2020). All authors have contributed equally.*

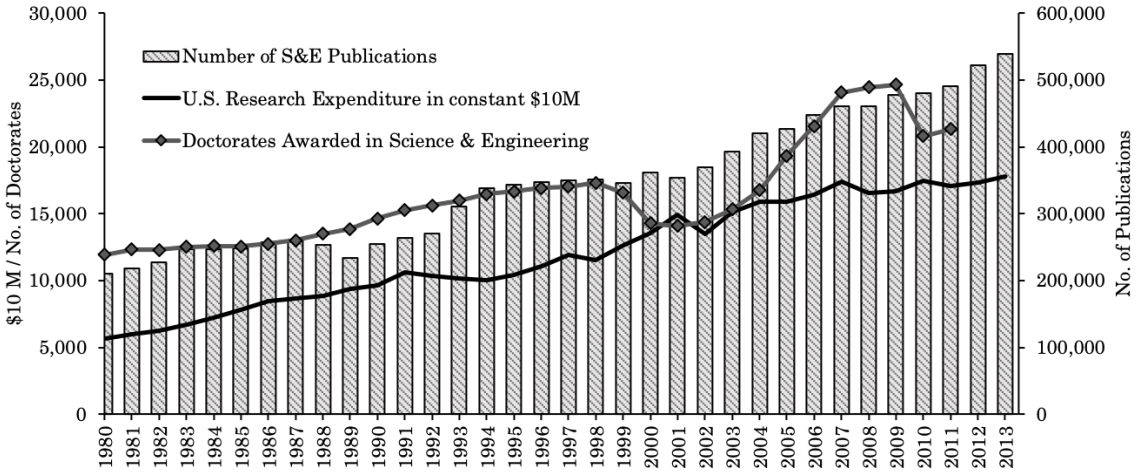


FIGURE 1.1: U.S. Scientific Investment and Output (1980-2013)[2]

tivity) grew substantially in the middle of the twentieth century, from 1.79 percent per year between 1870 and 1920 to 2.82 percent per year between 1920 and 1970. However, in the most recent period (1970-2014), productivity grew by a modest 1.62 percent per year. Gordon concludes that productivity rose between 1920 and 1970 largely because of significant technological progress, but more recently technical advance has been much less potent in spurring growth.

This slowdown is surprising given the sustained expansion of scientific input (measured in terms of research dollars spent) and output (measured by academic articles published) from American academia, as shown in figure 1.1. Indeed, Bloom et al. (2017) present evidence across a number of sectors showing that research productivity in the U.S. has declined since the 1970s. For instance, maintaining the exponential growth in semiconductor performance (otherwise known as "Moore's Law") in 2014

[2]Doctorates Awarded in S&E are calculated from the NSF's Survey of Earned Doctorates and excludes degrees in the Social sciences. Number of S&E Publications are from the Clarivate Web of Science and includes all scientific articles in the Science Citation Index-Expanded (SCI-EXPANDED) and Conference Proceedings Citation Index-Science (CPCSI-S) with a U.S. author from 1980 to 2013. U.S. Research Expenditure figures are calculated from the *National Patterns of R&D Resources: 2014-15 Data update. NSF 17-311.* tables and includes both basic and applied research expenditure. Figures are adjusted to 2016 dollars using GDP deflator from the World Bank National Accounts dataset.

required around 18 times the number of researchers it used to take in 1971. While growth rates for yields per acre for corns, soybeans, cotton, and wheat have averaged around 1.5 percent, the number of researchers in the agriculture sector has grown by a factor between 3 (wheat) and 25 (soybeans), a research productivity decline of about 4 to 6 percent per year. In the life sciences, the number of researchers has been rising by 6 percent annually, while research productivity measured by the discovery of new molecular entities per number of researchers has been falling by 3.5 percent per year.

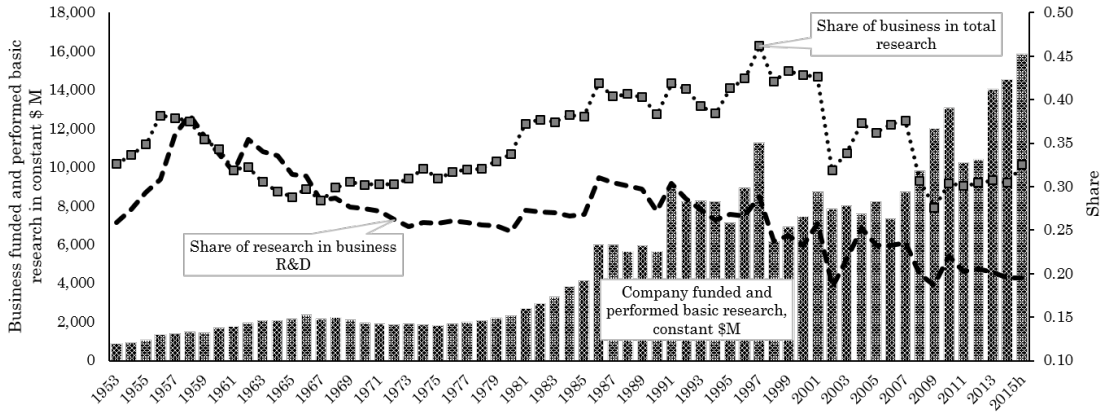


FIGURE 1.2: Business funded and performed research in the United States (1953-2015)[3]

Gordon attributes the rapid pace of technological progress in 1920-1970 to the development and extension of earlier fundamental technologies, such as the internal combustion engine and electricity. This process, which was often accompanied by important advances in science and engineering, was largely carried out by researchers working in corporate labs, which, by the 1920s, had replaced individual entrepreneurs as the primary source of American invention Gordon (2016, p.571-2).

[3]Data for this graph is sourced from the *National Patterns of R&D Resources: 2014-15 Data update. NSF 17-311.* from the National Science Foundation, National Center for Science and Engineering Statistics. 2017. Arlington, VA. Available at https://www.nsf.gov/statistics/2017/nsf17311/.

> "Much of the early development of the automobile culminating in the powerful Chevrolets and Buicks of 1940-41 was achieved at the GM corporate research labs. Similarly, much of the development of the electronic computer was carried out in the corporate laboratories of IBM, Bell Labs, and other large firms. The transistor, the fundamental building block of modern electronics and digital innovation, was invented by a team led by William Shockley at Bell Labs in late 1947. The corporate R&D division of IBM pioneered most of the advances of the mainframe computer era from 1950 to 1980. Improvements in consumer electric appliances occurred at large firms such as General Electric, General Motors and Whirlpool, while RCA led the early development of television."

However, this traditional innovation ecosystem dominated by the central research lab at large corporations began to be replaced by a more decentralized system featuring universities and VC-backed startups that sell their inventions through active markets for technology (MFTs). This change in the American innovation ecosystem is what motivates the dissertation. The following section briefly elaborates on these changes before outlining the three following chapters of the dissertation.

1.1 The Changing Structure of American Innovation

The new innovation ecosystem emerging since the 1980s is characterized by i) a decline of large corporate science ii) growing importance of academic and startup inventions iii) and the deepening of markets for technology.

1.1.1 Decline of Corporate Science

From the beginning of this era, large corporations began to look to universities and startups for ideas and new products. A good example is IBM, which on November 6, 1980 signed a contract with a then small firm, Microsoft, for the development of

its operating systems. Microsoft itself developed its operating system (eventually named the IBM PC-DOS) building on the operating system of another small company, Seattle Computer Products. As reliance on externally sourced inventions grew, many leading Western corporations also began to withdraw from scientific research (Mowery, 2009; Arora et al., 2018b).

Some corporate labs were shut down and others spun-off as independent entities. Bell Labs was separated from its parent company AT&T and placed under Lucent in 1996; Xerox PARC was also spun off into a separate company in 2002. Other labs were downsized: IBM under Louis Gerstner re-directed research toward more commercial applications in the mid-90s (Bhaskarabhatla and Hegde, 2014).[4] A more recent example is DuPont's closing of its Central Research & Development Lab in 2016. Established in 1903, DuPont research rivaled that of top academic chemistry departments. In the 1960s, DuPont's central R&D unit published more articles in the *Journal of the American Chemical Society* than MIT and Caltech combined. However, in the 1990s, DuPont's attitude toward research changed and after a gradual decline in scientific publications, the company's management closed its Central Research and Development Lab in 2016.[5]

NSF data indicate that the share of research (both basic and applied) in total business R&D in the U.S. fell from about 30 percent in 1985 to below 20 percent in 2015 (figure 1.2). The figure also shows that the absolute amount of research in industry, after increasing over the 1980s, barely grew over the 20 year period between 1990 to 2010. Other data show the same decline. Utilizing data on scientific publications, Arora et al. (2018b) show that the number of publications per firm fell

[4]According to personal communications with Ralph Gomory (former research director and Senior Vice President for Science & Technology at IBM), IBM even downplayed to investors the discovery of the scanning tunneling microscope (which earned Gerd Binnig and Heinrich Rohrer of the IBM Zurich Research Laboratory the Nobel prize in physics in 1986), for fear of a drop in share price.

[5]https://cen.acs.org/articles/94/i1/DuPont-Shutting-Central-Research.html

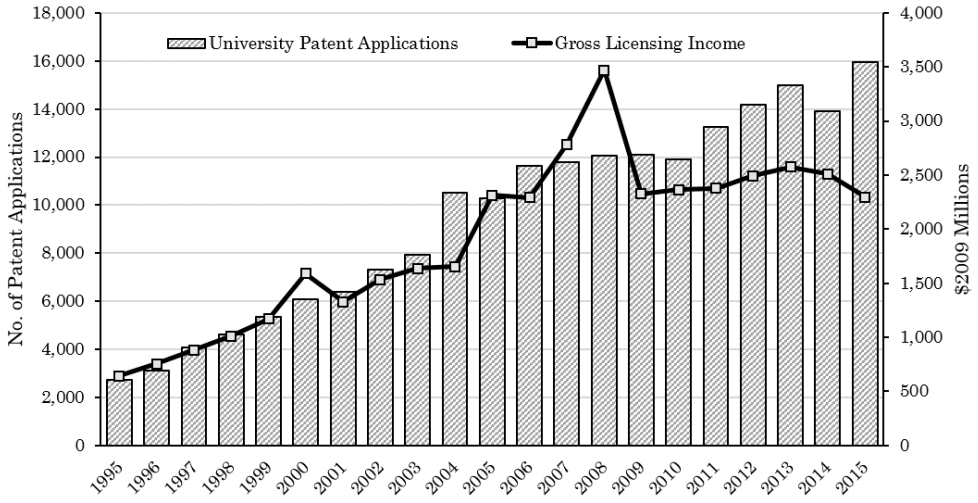


FIGURE 1.3: Patent Applications and Gross Licensing Income by Universities (1995-2015)[7]

at a rate of 20 percent per decade from 1980 to 2006 for R&D performing American listed firms.[6] Large firms' withdrawal from science can also be gleaned from the list of R&D 100 awards winners. Fortune 500 firms won 41 percent of the awards in 1971, but only 6 percent in 2006 (Block and Keller, 2009).

1.1.2 The Rise of Academic and Startup Inventions

In contrast, the research university sector continued to grow at a sustained pace from 1980 to 2016. Academic institutions spent $61 billion on basic and applied research in 2015. Their share of total research in 1985 was 23.8 percent and rose to 33.6 percent in 2015 (Boroush, 2017). Universities participated in the division of innovative labor by producing scientific insights, as well by directly producing inventions to be developed.

[6]The authors also find that the drop is even more dramatic for established firms in high quality journals. For articles within the top quartile of Journal Impact Factor scores, the magnitude of the drop increases to over 30 percent.

[7]This plot graphs university participation in technology markets using survey data from the Association of University Technology Managers (AUTM). The bar graph shows the number of patent applications filed by universities. The line graph shows gross licensing income received by universities. Units are in millions of 2009 dollars (deflated using GDP figures from Bureau of Economic Analysis, National Economic Accounts, Gross Domestic Product, http://www.bea.gov/national/)

The number of university patent applications more than quintupled between 1995 and 2015 from around 2,700 to over 15,000 per year. A similar trend is observed for university licensing income, which jumped from $.6 billion to $2.3 billion in the same period. University scientists have found it increasingly attractive to start their own businesses, with high-powered incentives and fast decision-making that are difficult to replicate in large, established firms. Changes in the institutional and legal environments complemented these trends. Start-ups can now get financing from venture capitalists and from SBIR and other government programs (Lerner, 2000; Mazzucato, 2015). Indeed, many firms have been spun-off from non-profit research institutions bringing forth such innovations as the MRI, recombinant hepatitis B vaccine, atomic-force microscope and the Google pagerank algorithm.

Moreover, recent years have seen the emergence of small, specialized research organizations that trade *ex ante* (research and consulting projects) and *ex post* (patents, software licenses, chip designs) knowledge products. These smaller firms either directly commercialize their ideas by introducing new products to the market or indirectly by selling them on to larger firms with downstream capabilities, in sharp contrast to the earlier system, where large firms originated their own inventions.

1.1.3 The Growth of Markets for Technology

American corporations reported $92 billion of income from licensing intellectual property in 2002, and the supporting IRS data show an annual growth of 11 percent from 1994 to 2004, which outpaced average GDP growth (3.42 %) in the same period (Robbins, 2009). The number of transferred patents as measured by reassignments between firms has also risen substantially from around 7,000 to over 12,000 cases per year between 1987 and 2014.[8] Moreover, business models specializing in selling

[8]Authors' calculations based on data from the USPTO Patent Assignment Dataset (Graham et al., 2018), replicating cleaning procedures in (Serrano, 2010) to identify patent reassignments that qualify as market transactions.

intellectual property without engaging in downstream manufacturing and sales have been validated by firms such as Exponent (chemicals), Genentech (biotech), and ARM (fabless semiconductor design).

The three trends surveyed above together point to the emergence of a new division of innovative labor, with universities focusing on research, large firms focusing on development and commercialization, and spinoffs, startups, and university technology licensing offices responsible for connecting the two.

Table 1.1: U.S. Distribution of technology licensing receipts by sector for 2002 & 2011 (in billions of dollars)

Sector	Licensing of IP Protected as Industrial Property (2002)	Technology Royalty and License Fee Income (2011)		
		Total	Tech and Ind Process	Software
Manufacturing	59.5	25.7	24.8	0.9
Wholesale, retail, transport	1	49.6	49.4	0.3
Information	1.9	27.7	2.1	25.6
Finance and Insurance	0.2	1.6	1.3	0.3
Professional and bus. services	3	4.5	2	2.5
Other industries	1	1.2	1.1	0.1
Total	66.6	111.2	81	30.2

Notes: This table shows the distribution of technology licensing receipts in the United States. The figures for 2002 are from Robbins (2009) Table 4.10. The figures for 2011 are from the Census Enterprise Statistics Program (ESP) 2011 Table 3: Royalty and License Fee Income from Rights to Use Intellectual Property (Detail). https://www.census.gov/econ/esp/historical.html

1.2 Outline of the Dissertation

Given the continued growth of scientific output in America that coincides with an expansion of markets for technology (MFTs), the second chapter of the dissertation, authored jointly with Ashish Arora and Sharon Belenzon, asks whether and how science and MFT may be related to each other. We theorize that science reduces transaction costs and increases gain from trade in inventions. The empirical prediction is that patents conceptualized in scientific principles are traded more often than those that are not. We test this in the context of the American patent market and confirm that patents that cite scientific articles and are closer to the scientific literature are more likely to be traded, even after accounting for differences in invention quality. The finding that science can increase the allocative efficiency of existing inventions is a novel mechanism that goes beyond the traditional role of science as a direct input into invention.

The advent of a new innovative ecosystem marked by a deepening division of innovative labor behooves us to ask whether it offers a better chance at hastening the pace of technological change than the preceding system of large corporate laboratories. Recent work shows that the exploitation of science for innovation is deeply uneven if not slow, with investments in information and communications technology taking the lion's share of VC investments, followed closely by biotechnology startups (Lerner and Nanda, 2020). This leaves innovations in "deep tech" or "tough tech" ranging from nuclear fusion, carbon capture and storage, advanced bioinformatics and optical computing, neglected (Arora et al., 2019b, 2022b). Moreover, corporate research has a number of characteristics that are hard for universities or startups to emulate: large corporations have access to complementary manufacturing resources; can more easily integrate multiple knowledge streams; and direct their research toward solving specific practical problems. This makes research at large corporations

more likely to produce commercial applications. Indeed, Bikard (2015) finds corporate publications to be 23 percent more likely to be cited than university publications one the same scientific discovery.

Given the enduring importance of corporate research in innovation, chapter 3 of the dissertation, coauthored with Ashish Arora, Sharon Belenzon, Konstantin Kosensko, and Yishay Yafeh, asks why scientific research at large American firms began. The earlier literature on the question has explained how the role of larger firm size and wider scope allows for the unpredictable payoffs of scientific research to be internalized by firms (Nelson, 1959a; Arrow, 1962). The subsequent literature has focused on the role of spillovers to rivals (Bloom et al., 2013; Arora et al., 2021b) and the role of scientific investments in absorbing such spillovers (Cohen and Levinthal, 1989). Chapter 3 builds on this literature and argues that a dearth of public science incentivizes firms at the technological frontier to invest in research. We test this idea on American firms during the early twentieth century when public science from American universities lagged behind European counterparts, especially in key areas such as organic chemistry and quantum physics. We find that firms such as Eastman Kodak, DuPont and General Electric that drew from these disciplines and reached the technological frontier were more likely to employ scientists and publish papers in scientific journals.

The fourth chapter of the dissertation asks why startups are more likely to innovate novel products using science than incumbent firms. Even as science reduces the costs of technological search (Mokyr, 2002; Fleming and Sorenson, 2004), firms may choose to exploit it to improve existing applications or invent solutions to new ones. The literature has generally found that startups tend to gravitate toward the latter while incumbents stick to incremental improvements (Tripsas, 1997; Mitchell, 1989). Inherent capability differences may explain this pattern: startups may be founded by scientists that are closer to the scientific discovery (Zucker et al., 2002) while in-

cumbents may suffer from organizational blinders that preclude them from jumping to new avenues of research (Sah and Stiglitz, 1986; Agrawal et al., 2010). On the other hand, however, startups may also select into more novel innovation even if they possess the same capabilities as incumbent firms. Testing this idea has been hitherto difficult because startup and incumbents often differ in their proximity to scientific discoveries and linking such discoveries to innovations (i.e., commercialized products) is difficult. I focus on the laser industry where scientific papers can be linked to products using their optical gain medium and exploit the influx of Soviet physics knowledge into America following the end of the Cold War to examine whether the startup propensity toward novel innovations is attributable to selection. I find that startups and incumbent firms do not differ in their pre-entry capabilities in absorbing new laser knowledge from the erstwhile Soviet Union, but that after entering the product market, startup products are more novel than incumbents'. Interestingly, startups are less likely to enter the product market conditional upon patenting but more likely to survive after entering the product market. I deduce that higher commercialization costs faced by startups incentivizes them to select into more novel innovations that offer higher payoffs at higher risk.

In summary, the three chapters constitute an effort to understand the mechanisms underlying the creation and exploitation of scientific knowledge by corporations.

2

Science and the Market for Technology

2.1 Introduction

A well-functioning Market for Technology (MFT) enhances welfare by allowing inventors to sell or license inventions to innovators who may commercialize them more efficiently.[1] The past several decades have witnessed a rise of a Market for Technology (MFT). Estimates based on Graham et al. (2018) show that patent reassignments have risen tenfold from around 2,000 to over 20,000 cases between 1980 and 2016. As well, U.S. corporations have reported a steady increase in royalty receipts and payments for industrial processes abroad, from $1.5 billion and $0.4 billion respectively in 1987 to $12.8 billion and $4.5 billion in 2017.[2] University licensing revenues have increased tenfold over an even shorter period, from $218 million in 1991 to $2.5 billion in 2015 (AUTM, 2015). At the same time, the reliance of inventions on science has increased, as indicated by the rise in the share of patents citing science from 4%

[1] *This chapter is adapted from joint work with Ashish Arora and Sharon Belenzon published in Management Science (Arora et al., 2022a). All authors have contributed equally.*

[2] Excludes receipts and payments from affiliates. Data for 2017 from BEA website (https://apps.bea.gov/iTable/iTable.cfm?reqid=62step=9isuri=1&product=4); data for 1987 from the scanned issues of the Survey of Current Business. 1921-2014. https://fraser.stlouisfed.org/title/46, accessed on March 11, 2019.

to 28% of all U.S. utility patents between 1980 and 2015 (Marx and Fuegi, 2020).[3] In this paper, we relate these two phenomena theoretically and empirically. Specifically, we investigate whether reliance on science increases the likelihood that the resulting invention will be traded by the inventor, and explore the possible channels for the relationship.

We define science as knowledge that is codified in the scientific and engineering literature. Inventions can be said to rely on science if they arise from scientific discoveries, if they draw upon scientific knowledge in important ways, or if they are conceptualized in scientific terms. Reliance on science may result in higher quality inventions, but may also increase the likelihood of trade by reducing transfer costs and increasing gains from trade. Science elucidates and codifies from empirical regularities the mechanisms underlying natural phenomena (Arora and Gambardella, 1994; Mokyr, 2002). Inventions conceptualized in scientific terms may be codified more effectively and their "metes and bounds" delineated more crisply, thereby reducing contracting frictions with potential buyers. Furthermore, codification in scientific terms may facilitate the discovery of new applications, more valuable than those contemplated by the original inventor, thereby raising potential gains from trade.

We develop a simple analytical framework which incorporates three mechanisms through which reliance on science could affect the probability of trade (MFT): invention quality, transaction costs, and gains from trade. In turn, these have separate implications for the relationship between reliance on science and probability of trade differ, by inventor size and by invention novelty. The market level equilibrium with

[3]This rise has coincided with an increase in the amount of scientific research produced each year. In 2016, 32,246 "hard science" doctorates were awarded in the United States, which is more than twice the number in 1986 (13,914) (Thurgood et al., 2006). "Hard science" includes Science and Engineering, excluding Social Sciences, Education, Humanities and Arts. Globally, the publication of peer-reviewed scientific articles has grown at an accelerating rate, with annual growth rates of 1.8% in the 1980s rising to around 4.01% in the 1990s and 3.99% in the 2000s. In aggregate, 1.7 million articles were published in 2016, compared to just over 500 thousand in 1980. (Authors' calculations based on Clarivate Web of Science.)

entry into invention shows that an increase in demand in MFT encourages the entry of smaller inventors. Insofar as reliance on science makes it easier to trade inventions, the entrants are more likely to rely on science for their inventions and to trade their inventions. This also highlights the empirical challenges in estimating the effect of reliance on science on MFT. Specifically, insofar as greater demand attracts specialized research organizations (e.g., universities and small firms) to supply inventions for trade, and insofar as we do not measure the commercialization capability of the inventor properly, this would lead to a positive association between the observed reliance on science in invention and the propensity to trade. More generally, unobserved differences in demand conditions in MFT may bias the estimated relationship away from the true one.

We mitigate this concern in several ways. First, at the patent level, we control for size using a variety of measures, and show that not only does the relationship continue to hold but that MFT-science relationship is stronger for small inventors, as predicted by our model. Second, we show that the MFT-science relationship is systematically stronger for novel inventions, making it unlikely that unobserved heterogeneity is the only cause of the relationship. Third, we aggregate up to the IPC level to study how the proportion of science-based patents is related to the proportion of patents that are traded. Consistent with our analytical framework, reliance on science is associated with entry of inventors, and a greater share of specialized inventors. However, the principal payoff from the IPC level analysis is that we can exploit changes in U.S. government funding of research following the collapse of the Soviet Union as a source of exogenous variation in the reliance on science. The identifying assumption is that the reliance on science is greater when the supply of relevant science increases for reasons unrelated to trade. We instrument for the share of inventions relying on science using changes in government funding to scientific fields relevant to the focal patent. We show that scientific output rises in response to government funding, which

in turn increases the share of science-based patents, and consequently, a higher rate of trade.

We use data on U.S. patents and scientific publications between 1980 and 2016. We measure reliance on science by whether a patent cites a scientific article (Marx and Fuegi, 2020). This dataset matches front page Non Patent Literature (NPL) citations in U.S. patents to peer-reviewed scientific publications from Microsoft Academic Graph (MAG). We use patent reassignments from the USPTO Patent Assignment Database (PAD) to measure transactions in MFT. We validate this as a measure of reliance on science in a variety of ways, and show further that our results are robust to using a textual similarity measure.

We present three main findings. First, a patent that cites a scientific publication has a 23% higher probability of being traded compared to a patent that does not cite science. After we control for quality using number of claims, triadic patenting status, (and stock market value for a subset of patents matched to listed firms), the magnitude of the science-MFT coefficient is reduced by a third, but still positive and statistically significant. We interpret this as suggesting that reliance on science is related to MFT in other ways, not just through quality.

Second, consistent with the view that science increases gains from trade, we find that the relationship between science and MFT is 8.6 times larger for small firms compared to large firms. Since larger firms already have high commercialization capabilities (and hence derive lower value from selling), reliance on science will have a smaller effect on the gains they can reap from selling their invention compared to smaller firms. We also find that the science-MFT relationship is about a third larger for patents using a one standard deviation "newer" combination of technological subclasses (Fleming, 2001). Unfamiliar inventions are likely to be more difficult for buyers to evaluate, but when such inventions are based in science, buyers may gain a better understanding of otherwise unfamiliar technological components. We also find

that reliance upon more recent science is more strongly associated with patent trade. This suggests that the principal channel through which science affects technology markets is not by reducing transfer costs, since mature science is more likely to provide a "common language" or background knowledge for market participants (Arora and Gambardella, 1994). Newer science, and more specialized science, on the other hand, is arguably more important in illuminating uncharted applications of inventions, thereby increasing gains from trade.

Third, we present instrumental variable results that are based on an exogenous change in the availability of science, which affects the prevalence of science-based inventions unrelated to trade conditions. After the fall of the Soviet Union, federal funding for research fell by almost 40% for Space and Defense between 1986 and 1992, whereas it doubled in Medical and Energy. These changes led to substantial variation in the supply of new knowledge, and therefore, variation in the reliance on science in different fields of invention. The geopolitical circumstances that precipitated the end of the Cold War are exogenous to MFT. We find that patent classes with a standard deviation increase in the share of science-based patents experience a 6.5% increase in the share of reassigned patents. Our IV estimates are 13% smaller than our OLS estimates, but still positive and statistically significant.

Our main contribution is to establish that reliance on science enhances technology trade. We build on the literature on MFT (Ziedonis, 2004; Serrano, 2010, 2018; Ma et al., 2017; Arque-Castells and Spulber, 2017; Arora and Fosfuri, 2000; Gans et al., 2002). Gans et al. (2008) and Galasso and Schankerman (2018) identify the effect of a reduction in transaction costs from the resolution of uncertainty around patent property rights (from USPTO allowance events and court decisions respectively) on MFT. With the exception of De Marco et al. (2017), the relationship between science and MFT has not been explored. We use exogenous variation in the availability of science to provide the first estimate of the causal impact of reliance on science on

MFT.

We also contribute to the literature that aims at quantifying the social returns to science (Griliches, 1957; Mansfield, 1980; Arora et al., 2021b). Our results suggest that science contributes to social welfare through a separate channel – the efficient allocation of inventions, and indirectly therefore, by promoting a division of innovative labor.[4] Jones (2009) argues that the growing "burden of knowledge" implies increasing individual specialization and greater need for cooperation. The market for technology is an important mechanism for facilitating such cooperation. To the best of our knowledge, this mechanism of the social returns to science has not been previously systematically explored empirically.[5]

A secondary contribution is to elucidate the different channels through which reliance of science is related to the market for technology. Inventions based on scientific discoveries may be higher in quality. However, there are other channels as well. Our empirical results suggest that the reliance on science in invention is also associated with lower transaction costs and higher gains from trade. By clarifying the underlying mechanisms of unfamiliar inventions, and using standard terms, the relevance of the patent may be more apparent to a wider set of buyers.[6] Put differently, scientific

[4]The evolution of Light Emitting Diode (LED) technology illustrates this point. Semiconductors that emit light were discovered as early as 1907, when Henry Round, a British radio engineer, observed a light yellow light emitting from his silicon carbide-based detector. However, the mechanisms behind this observation required a better understanding of quantum theory before the phenomenon could be applied more broadly. Therefore, early LED inventions were done in vertically integrated firms such as TI (infrared LED in 1961) and GE (red LED in the same year) (Sethi, 2013; Stevenson, 2009). Deeper scientific understanding allowed specialized firms such as Universal Display Corporation (UDC) to enter by licensing their intellectual property on dopants to incumbents. See for instance UDC and BASF's patent deal in IMSExpert. "$96M in OLED Patents, "Fruitful" Purchase for 2017". National Law Review. August 12, 2016 Friday. https://advance-lexis-com.proxy.lib.duke.edu/api/document?collection=newsid=urn:contentItem:5KFJ-DRC1-F03R-N0XF-00000-00context=1516831.

[5]Our paper is also related to the growing literature that examines the use of science in inventions using patent citations to science (Narin et al., 1997; Azoulay et al., 2015; Fleming et al., 2019; Agrawal and Henderson, 2002; Belenzon and Schankerman, 2009; Fleming and Sorenson, 2004; Ahmadpoor and Jones, 2017; Veugelers and Wang, 2019)

[6]The Bessemer process, for instance, diffused rapidly after its metallurgical properties were

inventions offer greater gains from trade. Realizing these gains from trade must contend with transaction costs, including the costs of transferring tacit know-how as well as opportunism by buyers or sellers. A fundamental problem of selling knowledge is that the bargaining process requires inevitably disclosing the "secret sauce" to the buyer (Arrow, 1962; Anton and Yao, 1994). A mirror problem for the buyer is that inventions often require complementary tacit know-how to exploit, and therefore require the active cooperation of the seller (Arora, 1995; Von Hippel, 1994; Polanyi, 1958; Kogut and Zander, 1992). Science can help codify a greater share of this tacit complementary knowledge and reduce the risk of hold-up and bargaining breakdowns (Galasso and Schankerman, 2014; Merges and Nelson, 1990). Scientific patents are "crisper", which helps potential buyers better understand what they are buying.[7] In sum, reliance on science reduces transaction costs because scientific inventions are easier to codify, and cheaper to transfer, and less vulnerable to contracting failures.

2.2 Science and the Precision of Inventive Claims

Some inventions are not amenable to scientific conceptualization (e.g., purely mechanical innovations such as the zipper fastener and Velcro). Others, such as the Bessemer steelmaking process, were empirical discoveries whose scientific basis would be clarified after invention.[8] Yet others, such as Giulio Natta's polypropylene patents, were conceptualized in scientific categories from the very beginning.

A canonical example from the invention of polypropylene illustrates that patents

sufficiently understood.

[7] A prime example is the chemical industry, where patents are more effective tools of protecting fruits of R&D thanks to their reliance on the explicit description of "formulae, reaction pathways and operating conditions" represented via Markush structures work better in protecting property rights (Arora and Fosfuri, 2000; Levin et al., 1987).

[8] As the first inexpensive industrial process for mass producing steel, the process revolutionized steelmaking. However, early licensees failed to reproduce his patents reliably. Subsequent metallurgists understood the chemical reaction underlying the process and identified phosphorous impurities as the cause of the erratic performance of the invention (Carnegie, 2006).

that solve a similar technical problem differ in the extent of their conceptualization in science. The 1963 patent by Giulio Natta and his colleagues at Montecatini on polypropylene, a chain of propylene molecules, strung together, with the side chain methyl molecules (CH3) appearing in a regular manner, was a landmark invention.[9] Polymers such as polypropylene exhibit very different strength and heat resistance depending on their steric structure (i.e., how the component molecules are arranged). Even when the propylene molecules are chained linearly, the side chains can "branch" out differently: in "atactic" polymers, this branching out is random, which leads to amorphous polymers (e.g., weaker but more elastic materials); in "isotactic" polymers, the branches occur on the same sides, resulting in stronger and more crystalline plastics, while in "syndiotactic" polymers, it occurs in alternating fashion. Giulio Natta's contribution to science was to distinguish between these three different structures based on X-ray diffraction techniques.[10] Natta's patent cites his research (Natta et al., 1955) and relates the physical properties of the new substance to its structure. Moreover, Natta separately patented a syndiotactic polypropylene as well.[11] That Natta separately patented these two different types of polypropylene illustrates how science helps clarify the "metes and bounds" of inventions.

A competing process to produce polypropylene was also patented in 1958 by Hogan and Banks at Phillips Petroleum (US2825721A).[12] The patent and other historical sources indicate that the inventors successfully obtained a high melting point, high molecular weight material by running propylene over a chromium oxide catalyst (McMillan, 1979). Although Hogan and Banks had, in fact, produced polypropylene, they did not fully understand this (nor the steric structures of these molecules that

[9] US3112300A "Isotactic polypropylene"

[10] Natta shared the 1963 Noble Prize in Chemistry, along with Germany's Karl Ziegler.

[11] US3258455A "Polypropylene having syndiotactic structure"

[12] "Polymers and production thereof"

Natta was uncovering during the 1950s). Thus, the invention is described in terms of the steps undertaken, not in terms of scientific principles and structures.[13] Indeed, after Natta's patent was published, Hogan amended an application for another polypropylene patent to cite Natta's paper (Natta et al., 1955).[14]

A more recent example is from hydraulic fracturing ("fracking"). Fracking involves drilling a hole into a bedrock formation and injecting fluids at high pressure to create artificial cracks through which hydrocarbons can be collected. One area of innovation has been in the choice of fluids used — firms have variously used foams, gels and water mixed with chemical additives ("slick water") — and in proppants (particles used to prop open the fissures created by the fluid).

Fracking fluids need to be viscous. Traditionally, this is accomplished by increasing polymer (e.g., guar gum-based chemicals) concentration. However, a major drawback from higher polymer concentration is that the chemicals themselves are expensive and difficult to separate from the pumped hydrocarbons. U.S. Patent 7012044B2, "Oil and gas recovery from subterranean formations" by BJ Services Company is conceptualized using principles of rheology (a branch of physics that studies the flow of solid and liquid materials). The invention consists of adding a negatively charged carboxyl group to the guar gum. The negative charge causes the molecules to repel each other, but because they are connected ("cross-linked") to each other, the result is an increase in viscosity at much lower concentrations. The invention draws on a PhD thesis in chemical engineering that deals with the rheology of gels (Kesavan, 1994). The patent is clearly specified: substituting a negatively charged caroboxyl group and cross-linking the molecules using known methods. We

[13]For instance, using a particular catalyst is primarily justified by lack of previous trials rather than any reference to the tacticity of the polypropylene: "Prior to this invention, the polymerization of olefins to tacky and solid polymers had not been catalyzed by a highly oxidized metal oxide as the essential catalyst ingredient... " (p.1, U.S. Patent 2825721)

[14]US4376851 "Solid polymers of olefins and production of such polymers"

should note that crisper patents are not necessarily more valuable (e.g., this patent was applied for in 1997 but totals only 16 forward citations). There may be other, more effective, means of enhancing the efficiency of fracking fluids.

Indeed, U.S. Patent 7971643B2[15] by Baker Hughes claims the use of "relatively lightweight and/or substantially neutrally buoyant particles as proppant material in hydraulic fracturing treatments and as particulate material in sand control methods such as gravel packing, frac pack treatments, etc." (2011, p.26) This invention does not cite a scientific article, and essentially provides a recipe for a more effective fracking fluid. It has received 148 patent citations (as of July of 2021) since its application in 2003.

Even inventions based on standard science may still be conceptualized in scientific terms. For instance, IBM's patent US10305921[16] combines anomaly detection (applying statistics to packet traffic on a network) with graph theory (to characterize the computer network) to develop a stochastic optimization model for deciding whether a node is malicious or benign. As such, this is the application of known methods in computer science. The patent cites a computer science paper on detecting malicious nodes using graphical inference methods (Manadhata et al., 2014).

Compare this to U.S. patent 10848517B1, granted in 2020 to Security Scorecard Inc., on a method of determining the security risk faced by an enterprise. It involves collecting data on the enterprise, including data on the packet traffic on its network, and correlating it with data from past security breaches. No scientific literature is cited, and the method for calculating security risk is vague. A potential buyer would need substantial additional know-how and operating data to make this patent

[15] "Methods and compositions of a storable relatively lightweight proppant slurry for hydraulic fracturing and gravel packing applications"

[16] "Network security apparatus and method of detecting malicious behavior in computer networks via cost-sensitive and connectivity constrained classification"

operational.[17]

The foregoing examples illustrate how using scientific categories to describe the invention clarifies what the invention is and its limits. Compare the invention of increasing the fluidity of fracking fluids by adding carboxyl groups to guar gum with the one using "lightweight, neutral buoyancy particles" as proppants in fracking fluids. The latter is vague (apparently walnut shells are also included), and the importance of the weight and buoyancy are not apparent. These examples also illustrate that sometimes the invention involves a scientific discovery (as was the case with polypropylene patents), and sometimes existing scientific techniques can be combined to solve a problem (as in the IBM networking security patent). Finally, these examples also illustrate that the economic value or quality of an invention is distinct from whether it is science-based. For instance, the IBM patent discussed above has received 18 citations, whereas the Security Scorecard patent has received 96 citations.

2.3 A Model of Trade in Patents

We present a simple model to clarify the relationship between science and MFT. There are I inventors. Each inventor is endowed with an invention, whose use of science is indexed by s. In subsection 2.3.4 we consider entry into invention but for now, I and s are given. The inventor can commercialize the invention herself and earn $y_i = q_i(x_i + \epsilon_i)$, where q_i is the quality of the i^{th} invention, and x_i is the expected value the inventor can extract, and ϵ_i represents the idiosyncratic component of value.

[17] Another example from history underlines this point. German chemical firms in the early twentieth century successfully prevented American firms from replicating their inventions even after the U.S. government mandated compulsory licensing of German patents under the Trading with the Enemies Act (TWEA) of 1917. Companies such as BASF disclosed the bare minimum for the Haber-Bosch process (critical for the generation of nitrogen for fertilizers), for instance, by withholding information about catalysts (Haynes, 1945, pp.86-7). American firms such as DuPont also expended substantial R&D dollars to learn by trial and error how to replicate German dye patents (Hounshell and Smith, 1988).

For the i^{th} invention, there are N_i firms that may buy it to commercialize it themselves, earning a payoff y_{ik} (and zero otherwise), where $y_{ik} = q_i(x_{ik} + \epsilon_{ik}) - \tau_i$.[18] Here, τ_i represents transaction costs, and x_{ik} represents the systematic component of the value the k^{th} potential buyer can extract from invention i, and ϵ_{ik} is the idiosyncratic component. Transaction costs have several components. These include contracting costs, such as legal fees, and trading frictions due to imperfect contracting, as well as the cost of transferring accompanying know-how. The primary determinant of x_{ik} is commercialization capability of the buyer, reflecting how well the potential buyer commercialize the invention, although it may also reflect ability to enforce patents (Galasso et al., 2013).

We assume efficient bargaining: If trade with at least one of the N_i potential buyers offers a surplus that is at least as great as the transaction costs, the invention will be traded. This assumption sweeps away considerations of asymmetric information, or how the joint surplus is divided between the buyer and seller.

2.3.1 Probability of Trade

The invention is not traded if $y_i \geqslant \max_k^{N_i}\{y_{ik}\} \iff x_i + \epsilon_i - \max_k^{N_i}\{x_{ik} + \epsilon_{ik}\} \geqslant -\frac{\tau_i}{q_i}$. We assume that ϵ_i and ϵ_{ik} are distributed *iid* as Type I extreme value distribution (Gumbell distribution).[19] The probability the invention is not traded, P = Pr(No Trade)

$$P(\text{No Trade}) = \frac{\exp(x_i + \frac{\tau_i}{q_i})}{\exp(x_i + \frac{\tau_i}{q_i}) + \sum_k^{N_i} \exp(x_{ik})} = \frac{1}{1 + \sum_k^{N_i} \exp(x_{ik} - x_i - \frac{\tau_i}{q_i})} \tag{2.1}$$

[18] We assume that the inventor and potential buyers are equidistant from each other in the product space, so that any rent dissipation is the same, regardless of who commercializes the invention. Hence the baseline payoffs can be normalized to zero for all.

[19] We normalize the scale parameter to unity and the location parameter to zero. This normalization eases notation and does not affect the results.

Equation 2.1 illustrates, as discussed earlier, that reliance on science can increase trade through gains from trade, $x_{ik} - x_i$, and through a reduction in transaction costs relative to the quality of the invention, $\frac{\tau_i}{q_i}$.[20] The combined effect is to increase the probability of trade. To minimise clutter, we feature the transaction cost channel in the algebra below. Formally,

Result 0 Science based inventions are more likely to be traded.

2.3.2 Role of Science

Inventor capabilities

Recall that x_i represents the ability of the inventor to derive value from her own invention. If we call "small" inventors as those with lower x_i, these inventors are more likely to sell.[21]

$$\frac{\partial P(\text{no trade})}{\partial x_i} = \frac{\exp(-x_i - \frac{\tau_i}{q_i} + \bar{x}_i)N_i}{(\exp(-x_i - \frac{\tau_i}{q_i} + \bar{x}_i)N_i + 1)^2} = \frac{a}{(a+1)^2} \geqslant 0, \quad \text{where}$$
$$a = N_i \exp(-x_i - \frac{\tau_i}{q_i} + \bar{x}_i) \tag{2.2}$$

Inventor capabilities and science

$$\frac{\partial^2 P(\text{no trade})}{\partial x_i \partial s} = \frac{1-a}{(a+1)^3}\frac{\partial a}{\partial s} \qquad \begin{cases} \leqslant 0 & \text{if } a > 1 \\ \geqslant 0 & \text{if } a < 1 \end{cases} \tag{2.3}$$

[20] Through out we normalize transaction costs by invention quality. The gains from trade channel is explicitly featured when considering novel inventions.

[21] For ease of exposition, we show the proof with homogeneous buyers, where $x_{ik} = \bar{x}_i \forall k$ where $\bar{x}_i$ is the value that the buyer can derive from the invention, the probability of No Trade can be rewritten as $\frac{1}{N_i \exp(\bar{x}_i - x_i - \frac{\tau_i}{q_i}) + 1} = \frac{1}{a+1}$.

The probability of no trade, $P(\text{no trade }) = \dfrac{1}{1+a} \approx 0.9$ in the sample.[22] Therefore, $a \approx \frac{1}{0.9} - 1 \leqslant 1$ so that the science will accentuate the increase in the probability of trade as firm size falls. Intuitively, smaller firms have greater net surplus from trade $\bar{x_i} - x_i - \tau$. Their probability of trade will be more responsive to a decrease in τ or an increase in $\bar{x_i}$, as long as the *pdf* of ϵ is monotonically increasing. This will typically be the case when the probability of trade is low. Formally,

Result 1 Smaller firms are more likely to trade their inventions than larger firms, and the difference increases for science-based inventions.

2.3.3 Novelty of Invention

Novel patents are likely to have greater variation across buyers in their ability to extract value. We show greater heterogeneity among buyers increases trade, and this effect is stronger for science based novel inventions. Novelty may also affect transaction cost, which we analyze later.

Buyer heterogeneity

We analyze how the no trade condition in equation 2.1 is affected by an increase in the heterogeneity of buyers. We compare two cases, one where all external buyers have the same $x_{ik} = \bar{x_i}$, and the other where external buyers vary but $\mathrm{E}_k[x_{ik}] = \bar{x_i}$, and we set $q_i = 1$ to ease notation.

$$\begin{gathered} P(\text{No Trade}|\text{heterog}) = \frac{1}{1+b} < P(\text{No trade}|\text{homog}) = \frac{1}{1+a}, \\ b = \sum_{k}^{N_i} \exp(x_{ik} - \tau - x_i) \end{gathered} \tag{2.4}$$

[22] Note that $\dfrac{\partial a}{\partial s} = -\dfrac{\partial \tau}{\partial s}\dfrac{1}{q_i}\exp(-x_i - \dfrac{\tau_i}{q_i} + \bar{x}_i) = -\dfrac{\partial \tau}{\partial s}\dfrac{1}{q_i}a \geqslant 0.$

where the inequality follows from recognizing that the exponential is a convex function and applying Jensen's Inequality, so that $b > a$. Further, greater heterogeneity among buyers in terms of commercialization capability, x_{ik}, reduces the probability of No Trade.

$$\left.\frac{\partial P(\text{No Trade})}{\partial s}\right|_{heterog} = \frac{\partial \tau}{\partial s}\frac{b}{(b+1)^2}, \quad \left.\frac{\partial P(\text{No Trade})}{\partial s}\right|_{homog} = \frac{\partial \tau}{\partial s}\frac{a}{(a+1)^2} \tag{2.5}$$

Because $1 > b > a$, the probability of trade increases faster with reliance on science when buyers are heterogeneous. That is

$$\left.\frac{\partial P(\text{Trade})}{\partial s}\right|_{heterog} - \left.\frac{\partial P(\text{Trade})}{\partial s}\right|_{homog} = \frac{\partial \tau}{\partial s}\left(\frac{a}{(1+a)^2} - \frac{b}{(1+b)^2}\right) \geqslant 0 \tag{2.6}$$

Result 2 The probability of trade increases with science faster when buyers are heterogeneous.

The foregoing argument also has implications for the relationship between novelty and reliance on science. Insofar as novelty implies greater heterogeneity in valuation, novel inventions are more likely to be traded than other inventions, and reliance on science enhances the effect of novelty. However, it is likely that novel inventions involve greater transaction costs. For instance, novel inventions are more likely to require the transfer of tacit knowledge for successful implementation. Therefore, buyer heterogeneity and higher transaction costs cut in opposite directions, and the net effect on the level of trade is ambiguous. Nonetheless, if reliance on science is more effective in reducing transaction costs for novel inventions, then even if the effect of novelty is ambiguous, we show that reliance on science will enhance the likelihood of trade for novel inventions.

Science is likely to be more effective in reducing transaction costs for novel inventions. The scientific basis of such inventions will make their potential value more apparent to buyers, and it should also reduce the tacit knowledge required for successful transfer. It is therefore plausible that the fall in transaction cost due to science is greater for novel inventions than for incremental inventions. Formally, letting σ represent novelty of invention, we assume that $\frac{\partial^2 \tau}{\partial \sigma \partial s} \leqslant 0$. Consider two inventions, $i\&j$, and let $\sigma_i > \sigma_j$, so that i is more novel than j. The probability of no-trade (normalizing $q_i = q_j = 1$) is given by $\frac{1}{1+b_m}, m = i, j$, where $b_m = \sum_k^N \exp(-x_m + x_{mk} - \tau(\sigma_m, s)$. Evaluated at the point where the probability of trade is the same, so that $b_j = b_i = b$, we have

$$\frac{\partial P_i(\text{no trade})}{\partial s} - \frac{\partial P_j(\text{no trade})}{\partial s} = \frac{b}{(1+b)^2}\left(\frac{\partial \tau(\sigma_i, s)}{\partial s} - \frac{\partial \tau(\sigma_j, s)}{\partial s}\right) \leqslant 0 \qquad (2.7)$$

Result 3 Reliance on science increases the probability of trade more for novel inventions than for incremental inventions.[23]

2.3.4 Equilibrium: Specialization and Division of Innovative Labor

We now relax the assumption that inventions are exogenously assigned, and analyze entry into invention. The prospect of being able to trade an invention increases its value. We show that an increase in the number of potential buyers would result in a higher share of inventions that rely on science.

Suppose the inventor captures a fraction λ of the surplus if the invention is traded. The expected payoff from invention is $\Pi = \lambda \ln(\exp(x_i) + N(\exp(\bar{x} - \tau(s)) - R,$

[23]Buyer heterogeneity is related to gains from trade. Result 3 shows that transaction cost reductions are more important when gains from trade are higher, pointing to the sense in which these are synergistic.

where R is the investment required to produce an invention.[24] The marginal inventor, who is indifferent between inventing and not inventing, is characterized by $x^*(s)$ such that

$$\Pi(x^*(s)) = \lambda \ln\left(\exp(x*) + N(\exp(\bar{x} - \tau(s))\right) - R = 0 \tag{2.8}$$

Equation 2.8 highlights the different ways in which reliance on science and MFT are related. First, as Adam Smith noted, trade encourage specialization: $\frac{\partial x^*}{\partial N} = -\frac{P(\text{trade})}{1 - P(\text{trade})} \leqslant 0$. Second, the equilibrium size of marginal inventor relying on science is smaller than of the marginal inventor not relying on science.

$$\frac{d\Pi(x^*)}{ds} = 0 \implies \frac{\partial x^*}{\partial s}\frac{\partial \tau}{\partial s} = \frac{P(\text{trade})}{1 - P(\text{trade})} \implies \frac{\partial x^*}{\partial s} \leqslant 0 \tag{2.9}$$

Third, reliance on science implies a direct increase in trade, as well indirectly because the marginal inventor relying on science is smaller, i.e., $\frac{dP(\text{no trade})}{ds} = \frac{a}{(1+a)^2}\left(\frac{\partial \tau}{\partial s} + \frac{\partial x*}{\partial s}\right) \leqslant 0$. The share of science-based inventions is correlated with the number of potential buyers, creating a potential source of bias. To see this, suppose that s is a binary variable, and the probability an invention uses science is p, which is the same for all inventions. Further, suppose inventor types, x, are distributed uniformly on the unit interval. The equilibrium entry condition (equation 2.8) implies that only inventions that satisfy $x > x^*(s)$ are realised. Let $x_1 = x^*(s = 1), x_0 = x^*(s = 0)$. Equation 2.9 implies that $x_1 < x_0$. The observed share of inventions, $\tilde{p}$, that use science is given by

$$\tilde{p} = \frac{p(1 - x_1)}{p(1 - x_1) + (1 - p)(1 - x_0))} > p \iff x_1 < x_0. \tag{2.10}$$

[24]For a proof, see Anderson et al. (1992), pages 60-61. For simplicity we assume that $\tau_i = \tau(s)$ so that transaction costs depend only the use of science but not on the identity of the inventor, and assume homogeneous buyers.

Equation 2.10 implies the observed share of science-based patents, $\tilde{p}$, is greater than p, the true share of science-based patents. The share of traded patents in equilibrium is

$$\begin{aligned} \text{share traded} \; &= \underbrace{\int_{x_0}^{1} \mathrm{P}(\text{trade}|x, s=0)dx}_{\alpha} \\ &+ p \underbrace{\left(\int_{x_1}^{1} \mathrm{P}(\text{trade}|x, s=1)dx - \int_{x_0}^{1} \mathrm{P}(\text{trade}|x, s=0)dx \right)}_{\beta} \end{aligned} \tag{2.11}$$

Equation 2.11 highlights the empirical challenge in estimating the structural relationship between the reliance on science and MFT, represented here by β, at the market level.[25] We observe $\tilde{p}$ rather than p. Unobserved factors, such as an increase in the number of potential buyers, N, will directly increase the probability of trade, as well as increase $\tilde{p}$. This will bias the OLS estimate upward. Patent level regressions may also not be free from bias if we imperfectly measure x_i, the commercialization capability of the inventor. Equation 2.9 implies that x_i is negatively related to the reliance on science, and also, by assumption, negatively related to the probability of trade.

To summarize, patent level estimations estimate the direct relationship (i.e., the reduction in transaction costs and increases in gains from trade), but not the indirect effects due to entry. Moreover, there is potential upward bias if commercialization capability is measured inadequately. At the market (IPC) level, estimates combine the direct and indirect (entry of specialized inventors), but there is potential upward bias because the observed reliance on science is measured with bias.

We present estimates at both the patent and the IPC level. In addition, we

[25]The structural relationship represented by β has both a direct component (represented by $\mathrm{P}(\text{trade}|x, s=1) - \mathrm{P}(\text{trade}|x, s=0)$, and an indirect component, represented by the differences between x_1 and x_0, as seen in the lower limits of the integrals. The bias is related to the indirect component, because $x_1 = x_0 \implies \tilde{p} = p$

develop a source of exogenous variation in p at the IPC level to purge $\tilde{p}$ of the bias. Concretely, suppose that $p = p(K), p' > 0$ where K is the stock of relevant science. That is, we assume that the share of science based inventions is increasing in the stock of available science. As discussed below, we use changes in government support for science that are unrelated to conditions in the market for technology, as a source of exogenous variation in K, and therefore, in p. The identifying assumption is that these changes in the government funding ΔG, and the resulting changes in the stock of knowledge, ΔK, are orthogonal to N, the unobserved number of potential buyers in the MFT.

2.4 Data

We combine data on patents and peer-reviewed scientific publications. Our patent data is from the 2016 publication of PatStat and encompasses around 5.2 million utility patents granted by the USPTO from 1980 to 2016. Patent reassignment (transaction date, identity of buyers and sellers) are from the USPTO Patent Assignment Database (PAD) (Graham et al., 2018), which records details on the transfer of ownership between patent assignees. To account for sample truncation, we limit our sample to patents granted on or before 2011 (for which we observe reassignments until 2015).[26] The final sample consists of about 3.9 million patents, of which 6.3% are reassigned at least once. We describe next the main steps taken to construct the sample and main variables.

Science-based inventions — We define science-based inventions as those that make at least one citation to a scientific article (Narin et al., 1997; Arora et al., 2021b; Roach and Cohen, 2013; Cotropia and Sampat, 2013). We use data from Marx and Fuegi (2020), which matches US patents to scientific publications in Microsoft Academic

[26] About 58% of reassignments are within five years after patent grant.

Graph (MAG) to identify pairs of citing patents and cited scientific publications (see appendix A.1.1 for more details). We identify 724,395 patents out of 3,883,777 that cite at least one scientific article in MAG on their front page.[27] However, there could be still be measurement error. Science-based patents may not cite science, and conversely, some citations to science may be "pro-forma", not really reflecting reliance on science. Accordingly, we also analyze market (patent class) level regressions because aggregation should reduce classical measurement error: the share of science-citing patents in a patent class is arguably a more reliable indicator of the extent to which inventions in that class rely on science. Our results are robust to an alternative measures of reliance on science, in-text citations, and the textual similarity between patents and scientific publications, as shown below.

Patent reassignments — We measure MFT by the patent reassignments in the USPTO Patent Assignment Dataset (PAD) from 1980 to 2015 for patents granted on or before 2011 (Marco et al., 2015).[28] The USPTO records transfers of ownership that occur between patent assignees. While the reporting of transfers is voluntary, firms that acquire patents have an incentive to report transfers, particularly in enforcing the acquired patents. We build on prior researchers, who have cleaned reassignments data to obtain those related to MFT transactions (Marco et al., 2015; Serrano, 2010).[29]

Invention quality — We use three methods to measure the quality of the patented invention. First, we use data from Patstat to count the number of *forward patent citations* a patent has received and normalize this by the average number of citations received by all patents in the focal patent's publication year. Our second measure

[27] See appendix A.1.1 for further details on variable construction.

[28] We focus on patent trade, which, unlike licensing, entail an exclusive transfer of property rights. Reliable large scale licensing data are not easily available, particularly for private firms.

[29] 6.3% of our sample's patents are reassigned at least once. See appendix A.1.1 for details on the cleaning procedure.

is whether the patent is a *triadic patent* i.e., is registered in the three largest patent jurisdictions - the European, Japanese, and U.S. patent offices (Dernis and Khan, 2004). That the same invention is patented in all three offices implies that the value to the inventor is high. Third, for firms listed in American stock exchanges, we use the *stock market valuation* of patents from Kogan et al. (2017), which uses the excess stock returns for patenting firm on the date of the patent's issuance date recorded in the USPTO official gazette. In addition, we use the *number of claims*, and the *length of the first claim* as other measures for the quality of the patent.

Invention novelty — We use two alternative measures of novelty. The first uses *patent textual similarity* to prior patents. Building on Arora et al. (2018a), for each focal invention, we calculate its textual similarity score for all previous patents (all USPTO patents with an earlier priority date than the focal invention). We normalize the proximity scores vector of the top 100 closest citation pairs for each focal patent by dividing each score by the corresponding maximum pairwise textual score for the focal patent. We average the standardized scores to derive a single textual proximity score for each focal invention. The second measure of novelty is the technology *combination familiarity* measure from Fleming (2001). We count the number of times the same combination of patent sub-classes had appeared before the focal patent's publication date. The assumption is that combinations of sub-classes that appear more often should be more familiar.[30]

Size — We measure the commercialization capability of an inventor by its declared size in USPTO maintenance fee payment records. Firms with less than 500 employees

[30]In our sample, the combination familiarity score ranges from 0 (first combination of its kind) to 174 (appeared 174 times before) with a mean of 76.8. In the regression analysis, we refine this measure so that the count exponentially decays with time at an annual rate of 18%. That is, a previous patent subclass combination from five years ago is weighted by $exp(-\frac{1}{5}) = 37\%$. This time decay allows for even old technological combinations to exhibit higher novelty if sufficient time passes by between patenting activities.

are classified as small, and pay 50% lower filing and maintenance fees. Second, we use initial patent assignee names matched to public company names in Compustat from the DISCERN project (Arora et al., 2021b), and use inclusion in Compustat as another indicator of size.

Buyers' heterogeneity — We measure buyer heterogeneity as the top four-assignee concentration ratio by patent class-years. The higher the share of patents assigned to the top four patentees in a patent class, the more unequal the distribution of valuations of inventions. We first extract the assignee names that are disambiguated in the HBS inventor dataset (Lai et al., 2011).[31] We then calculate a four-assignee concentration ratio by dividing the patent stock of the four most frequent assignees by the patent stock of all assignees in a 4-digit IPC-year.[32]

Marginal inventor characteristics — In empirical analysis at the patent class level, we proxy the capability of the marginal inventor by the average size of patent holders, the total number of unique sellers, and the share of entrants to patenting in a 4 digit IPC-year. Average size of patent holders is defined as patents granted to "small" assignees based on application and maintenance fee payment divided by patent stock in each patent 4-digit IPC-year. To identify the number of unique sellers, we cluster similar assignee names by using string distance measures. Assignee name pairs that are sufficiently similar to each other are then treated as a single name.[33] The share

[31] Available from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5F1RRI

[32] By constructing this measure, we implicitly assume that the assignees approximate the potential buyers in a technology market, and that 4-digit IPC classes are appropriate delineators of technology markets.

[33] This prevents misspellings or differences in legal nomenclature (Corp, Inc, Ltd etc.) from classifying a single assignee into two different entities. To a limited extent, this strategy also allows us to identify and unify technology licensing arms or divisions of companies, provided the name of the company is long enough. We define entrants as assignees that are patenting for the first time since the beginning of our sample in 1980.

of new entrants is defined as the number of entrants as a share of total assignees for a 4 digit IPC-year.

Table 2.1: Summary Statistics for Main Variables

	Obs.	Mean	Std. Dev.	10%	50%	90%
Patent Publication Year	3883777	1998.886	8.725	1986	2000	2010
Reassignment Dummy	3883777	0.063	0.243	0	0	0
Cite Science Dummy	3883777	0.179	0.383	0	0	1
ln(IPC Combination Familiarity+1)	3878063	1.198	1.694	0	0	4
Textual Similarity	2425203	0.224	0.130	0	0	0
Small Entity Dummy	3689237	0.226	0.418	0	0	1
Compustat Patent Dummy	3883777	0.265	0.441	0	0	1

Notes: Reassignment is a binary variable equal to one if the patent has ever been reassigned in the USPTO PAD dataset. *Cite Science* is equal to one if there has been a citation to Microsoft Academic Graph (MAG), and zero otherwise. *Combination Familiarity* of a patent is constructed by counting the number of times a patent's IPC sub-class combinations have appeared in the past (details in Fleming (2001)). *Small Entity* is equal to one if an assignee is classified as a small entity by section 41 of the U.S. patent act, and zero otherwise. *Compustat Patent* is equal to one if an initial assignee is matched to a Compustat firm, and zero otherwise.

2.5 Econometric Framework

For a given patent, we confirm that reliance on science increases the probability of trade, especially when the invention is novel, the inventor is small, and faces heterogeneous buyers. These relationships hold as we aggregate up to the IPC level. In addition, we confirm that reliance on science is associated with greater entry into invention, especially of small inventors. Finally, we develop sources of exogenous variation in the reliance on science at the market level to estimate the structural relationship between reliance on science and MFT.

2.5.1 Baseline Trade Equation (OLS)

We estimate a patent level equation for the likelihood a patent is traded :

$$Reassignment_i = \beta_1 s_i + \boldsymbol{Z_i'\gamma} + \boldsymbol{\xi_t} + \boldsymbol{\psi_c} + \upsilon_i \tag{2.12}$$

Reassignment is equal to one for if the patent is reassigned at least once during its term and zero otherwise.[34] Reliance on science, s_i, is equal to one for patents with at least one NPL citation to MAG and zero otherwise, and Z_i contains a variety of controls for quality. As well, we include complete set of dummies for the patent grant year (ξ_t) and its 4-digit IPC (ψ_c). υ_i is unobserved patent level characteristics. We expect $\hat{\beta}_1 > 0$.

We examine Results 1, 2, and 3 and interact s_i with the inventor size, market concentration, and invention novelty, respectively. Per Result 1, we expect the level relationship between size and patent reassignment to be negative, as science increases the gains from trade more for small inventors than large. Result 2 implies that the interaction term between reliance on science and market concentration should be positive, while Result 3 implies that the interaction term between science and novelty should be positive, as science reduces the higher transaction cost due to newness.

We argued that reliance on science can lower transfer costs as well as increase gains from trade. To disentangle the relative importance of these two mechanisms, we use the characteristics of the science being cited in our patents. For example, if transfer cost reduction (τ_i) is the sole mechanism, we may expect patents building on older science to be more likely to trade. Mature and established theories and empirical results are likely to have weathered more frequent and rigorous tests of validity (and attempts at falsification). However, for gains from trade ($\bar{x}_i - x_i$), we would expect patents using younger science to trade more often. Patents building on "cutting edge" science have greater uncertainty relating to their commercialization value.

We also explore whether patents that draw more specialized science are more likely to be traded compared to patents that are more narrowly specialized. Jones

[34] A patent that is reassigned multiple times gets the same *Reassignment* value of one as a patent that has been reassigned once.

(2009) has argued that scientific progress has increased the returns to specialization, necessitating greater collaboration among such specialists. Jones (2009) himself focused on collaboration in the production of knowledge. However, it is plausible that the application of knowledge may also require collaboration. MFT is an important mechanism for such vertical specialization and collaboration. In other words, inventions that draw upon specialized science probably require more follow-on invention, and therefore, offer greater gains from trade.

Entry of small inventors

To see whether reliance on science favors small inventors, we estimate a specification at 4-digit IPC c, publication year t level,

$$\begin{aligned} Share\ of\ Small_{ct} =& \beta_0 + \beta_1 Share\ of\ Science\ Citing\ Patents_{ct} \\ &+ \boldsymbol{Z'_{ct}\gamma} + \boldsymbol{\xi_t} + \boldsymbol{\psi_c} + \upsilon_{ct} \end{aligned} \tag{2.13}$$

We proxy the commercialization capability of the marginal seller by (i) the share of small patentees (ii) the number of sellers and (iii) entrant share, and regress these against the share of science citing patents out of all granted patents at the 4 digit IPC-year level. We control for forward patent citations, share of triadic patents, average number of claims, and average length of the first claims because technological advances may encourage the entry of new sellers. We also include IPC and year fixed effects to exclude the effect of any year or technology class-specific differences.

2.5.2 Instrumental Variable Strategy and Other Robustness Analyses

Measurement error

It is plausible that we measure the reliance on science with error. We directly probe the robustness of our measure in three ways. First, we add a dummy for whether the citations to the scientific article also appears in the body text of the patent, in addition to the front page NPL section. Second, we weight the front page citations to

science by the number of forward citations received from MAG articles by the cited science. Third, we calculate a measure of textual similarity between focal patent text to scientific articles published in the Web of Science, using data from Arora et al. (2018a).

Unobserved heterogeneity and instrument variable analysis

The association between reliance on science and MFT may also reflect other factors. Equation 2.11 showed that unobserved differences in the demand for inventions, N, would result in a spurious correlation between $\tilde{p}$, the observed share of science citing patents, and the share of patents that are traded.[35] To address this concern, we exploit a quasi-experiment where the cost of relying on science falls due to an exogenous rise in the relevant knowledge for some inventions, but not others. Specifically, we use the reallocation of federal funding for R&D around the end of the Cold War as a source of exogenous change in the availability of relevant scientific knowledge. We use the changes in the predicted stock of scientific knowledge as an instrument for the share of patents citing science to purge the effect of unobserved demand factors, as well as purge it of measurement error. This procedure allows us to estimate the causal effect of science on MFT.

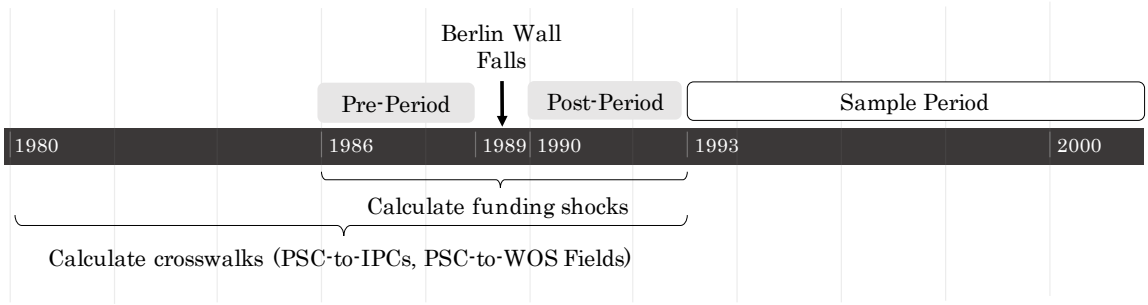


FIGURE 2.1: Timing Structure of Instrumental Variable Estimation

Government funding for R&D accounted for close to half of all research and devel-

[35] At the patent level as well, equation 2.9 implies that unobserved variation in commercialization capability is likely correlated with reliance on science, potentially creating biased estimates.

opment in the United States between 1980 and 1995 (see figure 2.2). Merrill (2018) shows that the end of the Cold War has resulted in a significant reallocation of the federal government's research portfolio, both within and between agencies. Between agencies, the dominant position of the Department of Defense (DoD) yielded to increased support for the Department of Health and Human Services. Within DoD, funding shifted out of physics, chemistry, and electrical engineering into computer sciences, oceanography, and biology during the 1990s. In a dataset of federal contract R&D data collected from USAspending.gov, we find that annual obligations of the U.S. government to contractors has decreased for defense and space technology, while increasing for medical and energy related items around the years after the end of the Cold War (see figure 2.2).[36]An increase in public funding for a scientific discipline increases the stock of knowledge, and hence increases the share of inventions relying on the focal science (Azoulay et al., 2018; Fleming et al., 2019; Moretti et al., 2019).

Our instrument is defined at the 4-digit IPC level and measure the i) the predicted number of relevant scientific papers to a 4-digit IPC and ii) the difference in average federal R&D funding between the post (1990-1992) and pre (1986-88) periods. Our first instrument is created in three steps. First, we create estimates of the stock of scientific knowledge (number of papers) relevant for each 4 digit IPC, using the share of NPL citations to different scientific disciplines by patents in an IPC. Second, we run an OLS regression at the Web of Science Field-year level to predict the number of papers published in a given field and year between 1992 and 2000. Third, we weight the predicted stock of knowledge due to the shocks to federal R&D funding in each scientific discipline by its relevance to a focal IPC to create the predicted shock to scientific knowledge for each IPC due to the end of the Cold War.[37]

[36]Appendix table A.2 in section A.1.2 summarizes this resource reallocation

[37]Please see appendix section A.1.2 for details.

[38]This figure plots the aggregate value of the contracts signed by the federal government for

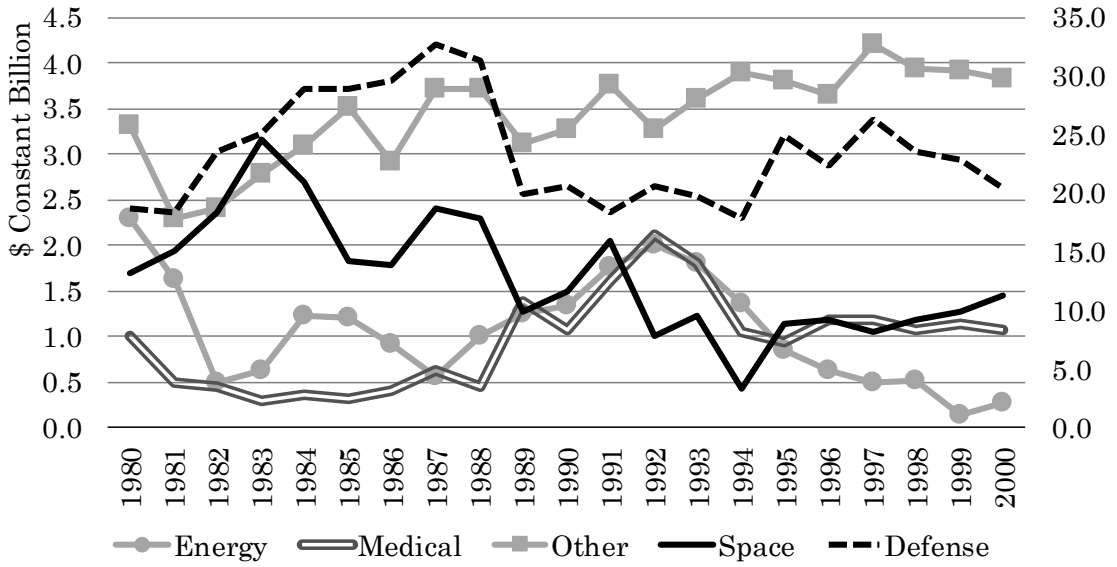


FIGURE 2.2: Government Contract R&D Funding, During and After the Cold War [38]

We implement our IV estimation with two-stage-least-squares. In the first stage, we predict the share of science-citing patents in a 4-digit IPC-year using our instrument, predicted relevant science. In the second stage, we estimate the share of traded patents as a function of the predicted share of science-citing patents. The first stage of this IV specification confirms that the reliance on science is a function of available knowledge stock, which we have shown in column 1 of table 2.10 to be positively related to post Cold War federal R&D funding shocks. The first stage regressions in table 2.10 shows that the relevant F statistics are over 100. Thus, changes in federal procurement spending due to the fall of the Soviet Union had an appreciable effect on the stock of scientific knowledge relevant to invention as well as on the share of patents relying on science.[39]

research and development, separated by main 2-digit Product and Services Codes: Energy (AG), Medical (AN), Space (AR), Defense (AC&AD). Figures are adjusted to 2012 dollars using GDP deflators from Louis Johnston and Samuel H. Williamson, "What Was the U.S. GDP Then?" MeasuringWorth, 2020.

[39]We allocate federal R&D funding to 198 Web of Science Fields by weighting the number of MAG publications matched to contracting DISCERN firms in each PSC. We then regress the

An alternate instrument is simply the difference in average government funding relevant to IPCs. For each fiscal year, we calculate the average federal R&D contract obligations for each Product and Services Code (PSC)[40] and match it to 4-digit IPCs based on the share of patents filed in each IPC by DISCERN firms contracting in the focal PSC, between 1980 and 1992.[41] We average these R&D dollars "relevant" to each 4-digit IPC, before (1986-88) and after (1990-92) the end of the Cold War. The difference of the natural logs of these values is our instrument that is used to predict the share of science-citing patents at the 4-digit IPC-year level after the Cold War (1992-2000).

A potential concern is that the end of the Cold War also changed demand conditions across technology fields. The federal government spent on average $315 billion a year on procurement between 1986 and 1992, where R&D contracts average $34.9 billion, accounting for 11% of total procurement. Specifically, non-R&D spending cuts may reduce demand in select fields, and thereby reduce the rate of invention in the field, possibly affect the share of science-based inventions as well as trade in technology. That is, changes in government procurement spending may also have affected N, the number of potential buyers of inventions (Lichtenberg, 1987).

We address this concern in three ways. First, the correlation between spending differences for R&D and non-R&D is positive but relatively low (r=0.257).[42] For

number of scientific publications with at least one author affiliation in the United States against the difference in logged funding for each Web of Science field between the pre (1986 to 1988) and post (1990 to 1992) period. Please see appendix A.1.2 for further details. In unreported robustness checks we verify that the results are unchanged when we weight the number of papers with forward citations received by the papers within five years of their publication..

[40] "The Product and Service Codes (PSC) Manual provides codes to describe products, services, and research and development (R&D) purchased by the federal government. These codes indicate "what" was bought for each contract action reported in the Federal Procurement Data System (FPDS)." 2015 Edition of the Federal Procurement Data System's Product and Service Codes Manual (Available from https://www.fpds.gov/downloads/top_requests/PSC_Manual_FY2016_Oct1_2015.pdf)

[41] Please see appendix section A.1.2 for details on the crosswalk.

[42] See figure A.3 for a visual comparison

instance, patent class "C07H" (SUGARS; DERIVATIVES THEREOF; NUCLEOSIDES; NUCLEOTIDES; NUCLEIC ACIDS) experiences a fourfold increase in its R&D funding between 1986 and 1992, while only a 3% decrease is observed for non-R&D funding. Pharmaceutical firms that receive medical research contracts also sell drugs (PSC 6505) and medical equipment (6515) to the federal government, which tends to be stickier (e.g. in VA medical centers). Second, we directly control for non-R&D federal contract spending. Third, we add robustness checks in appendix table A.3 that derives funding shocks net of number of patents, share of reassigned patents and patent forward citations.

2.6 Estimation Results

2.6.1 Reliance on Science and Trade

Table 2.2 contrasts the reassignment probability of patents that cite science and those that do not. We find that science-based patents are 1.4% (or 22% relative to the sample mean) more likely to be traded than those that are not based in science. In section 2.3, we argued that science leads to more trade because of its ability to reduce transfer costs (τ_i) and affect the gains from trade ($x_i - x_{ik}$). However, it is also possible that science-based inventions have higher quality (q_i). Indeed, table 2.2 shows that patents citing science have higher forward patent citations, are likely to be triadic patents, and have more claims and higher stock market values. Consistent with our model, reassigned patents also tend to exhibit higher values of these proxies of quality compared to those not reassigned.

Table 2.2: Mean Comparisons of Patent Characteristics, by Science and Reassignment Status

	T-Test		Cite Science = 0			Cite Science = 1		
	Diff.	Std. Error	Count	Mean	SD	Count	Mean	SD
Reassignment (%)	1.3688***	0.0321	3188613	6.037	23.817	695164	7.406	26.186
5-year Forward Patent Cites	3.3829***	0.0142	3188613	5.489	9.420	695164	8.872	15.482
Triadic Patent Dummy	0.1856***	0.0006	3188613	0.271	0.445	695164	0.457	0.498
Number of Claims	4.7607***	0.0166	3188154	14.485	11.505	695156	19.246	16.582
Mkt Value of Patent (KPSS)	5.7633***	0.0842	958233	11.384	35.552	259660	17.147	46.115
	T-Test		Reassignment = 0			Reassignment = 1		
	Diff.	Std. Error	Count	Mean	SD	Count	Mean	SD
5-year Forward Patent Cites	2.5321***	0.0226	3639800	5.935	10.422	243977	8.467	15.592
Triadic Patent Dummy	0.0465***	0.0010	3639800	0.301	0.459	243977	0.348	0.476
Number of Claims	2.3334***	0.0265	3639335	15.191	12.497	243975	17.524	15.217
Mkt Value of Patent (KPSS)	-1.5637***	0.1576	1156249	12.692	38.362	61644	11.128	33.312

Notes: Comparison of means at the patent level. Standard errors are in parentheses. Significance levels are annotated as * $< 10\%$ ** $< 5\%$ *** $< 1\%$

2.6.2 Baseline Trade Equation

Table 2.3 presents the Linear Probability Model (LPM) estimates. Controlling for year and 4-digit IPC fixed effects, column 1 shows that citing a scientific article is associated with a 23% higher probability of a patent being traded, relative to the sample mean.[43] Patents that cite science may represent higher quality inventions and better crafted patents. We therefore control for triadic patent status, number of claims, and the length of the first claim. More independent claims and shorter independent claims are related to broader patent scope (Kuhn and Thompson, 2019). In addition we measure whether a patent has been filed in multiple patent jurisdictions (U.S., Europe, and Japan). Such triadic patents tend to be of high private value, and hence, tend to be of higher quality.[44] Column 2 shows that number of claims and triadic patenting status are positively correlated with patent trade. The coefficient on science citation decreases in magnitude by 46%, but remains positive and statistically significant. For a subset of patents that are issued to U.S. listed firms, we measure their market valuations based on excess stock price returns of inventing firms on their grant dates (Kogan et al., 2017). Citing science continues to have a positive and statistically significant relationship with reassignment.[45]

Splitting the sample by technology classes (Columns 4-7, Table 2.3), we find that the *Cite Science Dummy* coefficient in the Life Sciences is around 55% as large as in ICT. Value chains in the ICT sector tends to be more complex, possibly muting the gains from having a clearer scientific grounding for one invention. The life sciences,

[43]Unless stated otherwise, the percentage magnitudes reported here are relative to the sample mean in each specification.

[44]In columns 2 and 3 of appendix table A.4, we also add dummies for quintiles and deciles of five-year forward patent citations. More highly cited patents are traded more, but the positive relationship between reassignment and science citation remains significant.

[45]Interestingly, the relationship between stock market valuations and patent trade is negative, perhaps because this captures the value idiosyncratic to a firm (x_i in the model) rather than common quality (q_i).

in contrast, have clearly delineated targets and therapeutic areas that are tackled by clearly structured molecular compounds. Therefore, transfer cost reduction from clarifying the underlying mechanisms by referencing science in an invention may be larger for the life sciences.

Table 2.3: Markets for Technology and Reliance on Science (OLS)

	Baseline Science			By Technological Sector			
	(1) Baseline	(2) Quality Controls	(3) Compustat Sample	(4) Life Sci	(5) Chem	(6) ICT	(7) Other
Cite Science Dummy	1.424**	0.768**	0.262*	0.898**	0.660**	0.578**	0.825**
	(0.038)	(0.039)	(0.130)	(0.122)	(0.092)	(0.058)	(0.074)
Forward Patent Cites		0.790**	0.354**	0.551**	0.786**	0.792**	1.000**
		(0.019)	(0.050)	(0.030)	(0.046)	(0.034)	(0.025)
Triadic Patent Dummy		0.806**	0.459**	2.305**	0.229**	1.092**	0.522**
		(0.029)	(0.160)	(0.100)	(0.077)	(0.058)	(0.041)
Number of Claims		0.067**	0.014**	0.069**	0.058**	0.050**	0.080**
		(0.001)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)
Length of First Claim		-0.000**	-0.000	-0.000**	-0.000	-0.000*	-0.000**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mkt Value of Patent (KPSS)			-0.002				
			(0.001)				
Avg of DV	6.283	6.283	5.589	8.680	6.564	5.678	6.099
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	No	No	Yes	No	No	No	No
R^2	0.010	0.014	0.157	0.012	0.010	0.018	0.013
N	3,883,776	3,882,632	797,914	357,124	473,782	1,020,997	2,030,724

Notes: Unit of analysis is the patent. Dependent variable is equal to 100 if the patent is reassigned and zero if not. Patent reassignment is a binary variable equal to 100 if the patent has ever been reassigned in the USPTO PAD dataset and zero otherwise. *Cite Science Dummy* is one if there has been a citation to Microsoft Academic Graph (MAG), and zero otherwise. *Triadic Patent Dummy* is one if the patent shares a prior art in the USPTO, EPO, and JPO, and zero otherwise. *Number of Claims* counts the number of independent and dependent claims in a patent. *Length of First Claim* counts the number of words in the first claim of the patent. *Mkt Value of Patent (KPSS)* is the value of a patent (in million dollars) based on the cumulative abnormal returns in the firm's market value at the issuance event of the patent Kogan et al. (2017). Standard errors are robust to arbitrary heteroscedasticity for all columns except for column 3, whose standard errors are clustered at the Compustat firm level.

Measurement error and validation

We introduce three robustness checks per section 2.5.2. We include "in-text" citations to MAG publications that appear in the body text of the patent. In-text citations do not affect the patentability of an invention as much as front page NPL citations. Therefore, their inclusion may signal a greater reliance on science by the inventor. We find in column 1 of table 2.4 that in-text citations to science are positively correlated with patent trade (though the coefficient is not statistically significant).

Table 2.4: Markets for Technology and Reliance on Science, Alternative Measures (OLS)

	DV: Reassignment=1		
	(1)	(2)	(3)
	In-text	Citation Weighted	Textual Similarity
Cite Science Dummy	0.760**		
	(0.042)		
Cite Science: Intext	0.024		
	(0.062)		
ln(Cite Science: FwdCitation Weighted)		0.122**	
		(0.006)	
Textual Similarity to Science			0.094**
			(0.007)
Forward Patent Cites	0.790**	0.788**	0.798**
	(0.008)	(0.019)	(0.019)
Triadic Patent Dummy	0.806**	0.805**	0.849**
	(0.028)	(0.029)	(0.029)
Number of Claims	0.067**	0.067**	0.069**
	(0.001)	(0.001)	(0.001)
Length of First Claim	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
Avg of DV	6.283	6.283	6.283
4-digit IPC Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Firm Fixed Effects	No	No	No
R^2	0.014	0.014	0.014
N	3,882,632	3,882,632	3,882,632

Notes: Unit of analysis is at the patent level. "Cite Science Dummy: In-text" is equal to one when a scientific article is cited in the body of the patent text as an in-text citation. "ln(Cite Science: FwdCitation Weighted)" takes the natural log of the total number of citations papers cited by a patent receive from other scientific papers in Microsoft Academic Graph. "Textual Similarity to Science" measure the textual overlap of a patent's text with abstracts of articles in Clarivate Web of Science's Science Citation Index - Expanded. All other variables definitions are identical to table 2.3. Standard errors are robust to arbitrary heteroscedasticity.

On the other hand, the science surrounding an invention may be too established or canonical such that too few citations are made. For example, GPS technology relies on Einstein's general theory of relativity for its time dilation corrections between satellites and ground receivers, but few GPS patents cite his papers. We therefore weight the science citation dummy by the quality of the cited science in column 2 of table 2.4 and find that patent reassignment is still positively correlated with the number of citations a cited paper receives from other papers within MAG.

We offer three additional validity checks of our citation-based measure of reliance on science. First, Arora et al. (2021b) show that patents that cite scientific publications are more likely to respond in the Carnegie Mellon Survey that public research findings (from government and academia) are important for their inventions. They also find that firms tend to recognize the importance of the field they cite in their patents. Firms that cite scientific articles in patents also tend to operate a greater basic scientific research program as a share of total R&D.

Table 2.5: Reliance on Science, for Inductees of the National Inventors Hall of Fame

Group		Variable	Obs.	Mean	Std. Dev.	Median
Patents by National Inventor Hall of Fame (NIHF) Inductees		Cite Science Dummy	97	0.567	0.498	1
		Textual Simil. to Science	97	4.634	2.002	4.875
	also Nobel Laureate	Cite Science Dummy	8	0.750	0.463	1
		Textual Simil. to Science	8	6.539	1.473	6.228
	also Lasker Laureate	Cite Science Dummy	1	1	.	1
		Textual Simil. to Science	1	9.183	.	9.183
	also Turing Laureate	Cite Science Dummy	4	1	0	1
		Textual Simil. to Science	4	4.292	2.126	4.718
	also Franklin Institute Award Laureate	Cite Science Dummy	11	0.636	0.505	1
		Textual Simil. to Science	11	5.546	1.656	6.054
	also NMTI Laureate	Cite Science Dummy	18	0.556	0.511	1
		Textual Simil. to Science	18	4.482	2.244	5.454
non-NIHF Patents		Cite Science Dummy	3,883,680	0.179	0.383	0
		Textual Simil. to Science	3,883,680	3.541	2.037	3.555

Notes: This table shows summary statistics on reliance on science for patents published between 1980 and 2011 that are also matched to inductees of the National Inventors Hall of Fame (NIHF). Nobel Laureates include those in physics, physiology or medicine, and chemistry only. The NMTI refers to the National Medal of Technology and Innovation. Cite Science Dummy is equal to one if the patent cites at least one scientific article from MAG. Textual Similarity to Science is a continuous measure of textual overlap between a patent and scientific articles from Clarivate Web of Science between 1990 and 2015.

Second, we focus on a subset of patents whose inventors were inducted to the National Inventor Hall of Fame (NIHF) and examine whether high-caliber inventors that are also scientists are more likely to cite science, compared to similarly high-caliber inventors. The NIHF inductees are recognized for "great technological advances that make human, social and economic progress possible," which may range from Post-it Notes to blue LEDs.[46] The NIHF lists descriptions of the nature of the invention and the contribution of the inductee, together with the USPTO patent number for the most representative invention of that inductee.[47] We are able to link 97 patents for 111 inductees for our sample period.[48] We manually searched for whether each inductee had been awarded any one of four prestigious awards in science: the Nobel Awards in physics, medicine, and chemistry, the Lasker Award (for medicine), the Turing Award (for Computer Science), and the Franklin Institute Award (Arts et al., 2020). We find that 57% of the patents by NIHF inductees cite science, which is more than three times the sample average. However, we find that this probability increases to 75% for patents by Nobel Laureates, while Lasker and Turing Award winners always cite science in their most representative patents listed by the NIHF. We also link these inductees to the winners of National Medals of Technology and Innovation (which does not specifically require a scientific contribution) and find that they cite science slightly less often (56%).

Third, we calculate an alternative measure of reliance on science based on the textual overlap of patent text with abstracts from scientific articles. We leverage data from Arora et al. (2018a), which calculates a weighted cosine similarity measure between the full text of patents in our sample and the abstracts of scientific

[46]Spencer Silver, Patent Number: US3691140A, and Shuji Nakamura, Patent Number: US5290393A.

[47]https://www.invent.org/NIHF-hall-of-fame-inductees-list-alphabetical

[48]While the inventions are patented between 1980 and 2011 to fit with our sample, inventor years of birth range between 1914 and 1975.

articles in Clarivate Web of Science's Science Citation Index - Expanded. Using this textual overlap measure between patent-paper pairs, we keep the top 100 patents most similar to the focal paper. We count the number of times a given patent is classified within this "top 100" set, which we use as an alternative measure for reliance on science. We also tried the patent level averages of the ranks, as well as the similarity scores with respect to scientific publications. We replicate the baseline results using this new measure in column 3 of table 2.4, finding that a standard deviation increase in textual similarity to science is associated with a 3% greater reassignment probability.

Table 2.6: Science and MFT, by Seller Size and Buyer Heterogeneity

	Seller Size		Buyer Hetero
	(1) Maint.Fee	(2) Compustat	(3) C4 Share
Cite Science Dummy	0.334**	-0.669**	0.538**
	(0.040)	(0.053)	(0.068)
Small Entity Dummy	0.073		
	(0.041)		
Small Entity Dummy × Cite Science Dummy	2.549**		
	(0.101)		
Non Compustat Dummy		0.890**	
		(0.033)	
Non Compustat Dummy × Cite Science Dummy		2.381**	
		(0.069)	
C4 Share			0.197
			(0.255)
C4 Share × Cite Science Dummy			1.355*
			(0.536)
Forward Patent Cites	0.814**	0.806**	0.919**
	(0.020)	(0.019)	(0.026)
Triadic Patent Dummy	0.887**	0.787**	-0.016
	(0.030)	(0.029)	(0.036)
Number of Claims	0.066**	0.070**	0.076**
	(0.001)	(0.001)	(0.002)
Length of First Claim	-0.000**	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)
Avg of DV	6.265	6.283	8.114
4-digit IPC Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
$R\hat{2}$	0.015	0.015	0.020
N	3,689,008	3,882,632	2,782,627

Notes: Unit of analysis is at patent level. Dependent variable is equal to 100 if the patent is reassigned and zero if not. *Small Entity Dummy* is equal to one if an assignee is classified as a small entity by section 41 of the U.S. patent act, and zero otherwise. *Non Compustat Dummy* is equal to one if an initial assignee is not matched to a Compustat firm, and zero otherwise.

Inventor commercialization capability

We investigate the empirical support for Result 1 that patents owned by smaller firms are more likely to be traded, especially for patents that cite science. We measure seller commercialization capability (x_i) through patent ownership by "small" firms recorded in USPTO maintenance fee payments, and by Compustat companies. We expect that smaller patentees are more likely to sell their invention than larger ones, and the association is stronger for patents that cite science. Table 2.6 column 1 shows that science-citing patents owned by small entities are around 7.6 times more likely to be sold than other patents citing science.[49] Column 2 shows that patents owned by non-Compustat owners are 14% more likely to be sold relative to the sample mean, while the gap with Compustat owners widens by 3.7 times for patents citing science. We replicate this result using textual similarity measures to science in columns 1 and 2 of table 2.8. The predicted reassignment gap between small and large firms is around 67% larger for patents in the 75th percentile of similarity scores compared those in the 50th percentile. The gap between non-Compustat and Compustat firms is around 35% larger for 75th percentile similarity scores compared to 50th percentile.

Buyer heterogeneity

Result 2 predicts that industries with more heterogeneous buyers will exhibit more trade, and reliance on science magnifies the gap. We measure industry concentration by the share of C4 patentee-owned patents in a focal patent's 4-digit IPC. Because C4 patentees are likely to be large entities, simply regressing reassignment against the share of C4 patentees will measure the effects of patentee size rather than buyer heterogeneity. Therefore, we limit our sample to patents owned by non-C4 paten-

[49] In unreported checks, we confirm that the level effect of regressing patent reassignment against small *Small Entity Dummy* without interactions with *Cite Science Dummy* in column 1 is positive and significant.

tees. Column 3 of table 2.6 tests these predictions. For non-C4 patentees, being in a concentrated market where 90% of all patentees are C4 patentees leads to a 4% gain in reassignment probability compared to one where only 10% of all patentees are C4 patentees. While citing science is associated with a positive increase in reassignment probability, the patents in concentrated (90% C4 share) markets experience an increase 2.6 times larger than those in unconcentrated (10% C4 share) markets. We also replicate these findings in column 3 of table 2.8 replacing citations to science with textual similarity to science.

Table 2.7: Science and MFT, by Patent Novelty (OLS)

	Patent Novelty		Science Characteristics	
	(1) ABCL	(2) Fleming	(3) Recency	(4) Specialization
Cite Science Dummy	1.339**	1.053**		
	(0.080)	(0.046)		
Avg(Lag to Cited Science)			-0.014**	
			(0.004)	
1-Normalized Field Counts				0.480**
				(0.103)
Textual Similarity	-1.923**			
	(0.133)			
Textual Similarity × Cite Science Dummy	-1.756**			
	(0.258)			
ln(IPC Combination Familiarity+1)		0.070**		
		(0.008)		
ln(IPC Combination Familiarity+1) × Cite Science Dummy		-0.223**		
		(0.017)		
Forward Patent Cites	0.611**	0.796**	0.602**	0.600**
	(0.020)	(0.018)	(0.029)	(0.029)
Triadic Patent Dummy	1.073**	0.825**	1.692**	1.675**
	(0.034)	(0.030)	(0.062)	(0.062)
Number of Claims	0.044**	0.067**	0.049**	0.048**
	(0.001)	(0.001)	(0.002)	(0.002)
Length of First Claim	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Avg of DV	4.877	6.274	7.401	7.420
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
$R\hat{2}$	0.013	0.014	0.018	0.018
N	2,424,715	3,877,819	830,868	830,868

Notes: Unit of analysis is at patent level. Dependent variable is equal to 100 if the patent is reassigned and zero if not. *ABCL (Textual Similarity)* averages the patent-to-patent pairwise text similarity scores from Arora et al. (2018a) for each focal patent, normalized by its maximum score. *Combination Familiarity* of a patent is constructed by counting the number of times a patent's IPC sub-class combinations have appeared in the past (details in Fleming (2001)). *Avg(Lag to Cited Science)* is defined as the average difference in the grant year of a patent and the publication year of a cited scientific article. The rest of the variable definitions are identical as table 2.3. Standard errors are robust to arbitrary heteroscedasticity.

Invention novelty

Recall that more novel inventions may have higher gains from trade but also higher transaction costs, implying that the relationship between novelty and trade is theoretically ambiguous. Result 3 implies, however, that novel patents that rely on science have a higher probability of trade than novel patents that do not cite science. We test Result 3 by interacting science with two measures of invention novelty:

textual similarity to prior patent art (Arora et al., 2018a) and patent subclass Combination Familiarity (Fleming, 2001). The interaction terms in columns 1 and 2 of table 2.7 are negative and statistically significant, consistent with our prediction. While both patents whose subclass combinations are in the 25th percentile of Combination Familiarity scores (in other words, novel patents) and those in the 75th percentile (not novel patents) exhibit positive association between citing science and patent reassignment, this association is 72% stronger for novel patents.[50]

Cited science recency and specialization

The empirical results thus far indicate that reliance on science is associated with greater trade, even after controlling for the quality of the patented invention. This suggests that reliance on science may also lower transfer costs and increase gains from trade. Patents using recent science may be more likely to trade if gains from trade are at work, but if patents citing older, "textbook", science are more likely to be traded, then reductions in transfer costs may be at work. Column 3 of table 2.7 tests these predictions on 830,868 patents that cite at least one scientific article in MAG and finds that patents citing more recent science are more likely to be traded. A one standard deviation decrease in the average citation lag (7.7 years) is associated with about a 1.5% higher trade probability relative to the sample mean. Column 4 explores whether patents drawing on specialized science are more likely to be traded. We measure specialization by one minus the number of Web of Science fields found in the NPL citations of a patent divided by the number of papers cited. We find that a patent citing a standard deviation increase in specialization (.320) is associated with a 2% higher probability of trade relative to the sample mean. This supports the view that science increases gains from trade.

[50]We confirm in unreported robustness checks that similar results hold with other text-based patent similarity measures (Kuhn, 2016).

Table 2.8: Textual Similarity to Science and MFT, by Seller Size, Buyer Heterogeneity, and Patent Novelty (OLS)

	Seller Size		Buyer Hetero	Patent Novelty	
	(1) Maint.Fee	(2) Compustat	(3) C4 Share	(4) ABCL	(5) Fleming
Textual Similarity to Science	0.074**	-0.092**	0.097**	0.149**	0.160**
	(0.007)	(0.012)	(0.013)	(0.014)	(0.008)
Small Entity Dummy	-0.296**				
	(0.066)				
Small Entity Dummy × Textual Similarity to Science	0.217**				
	(0.015)				
Non Compustat Dummy		0.170**			
		(0.059)			
Non Compustat Dummy × Textual Similarity to Science		0.294**			
		(0.014)			
C4 Share			-0.804*		
			(0.375)		
C4 Share × Textual Similarity to Science			0.231**		
			(0.088)		
Textual Similarity				-2.168**	
				(0.209)	
Textual Similarity × Textual Similarity to Science				-0.057	
				(0.050)	
ln(IPC Combination Familiarity+1)					0.149**
					(0.013)
ln(IPC Combination Familiarity+1) × Textual Similarity to Science					-0.038**
					(0.003)
Triadic Patent Dummy	1.056**	0.968**	0.159**	1.244**	0.990**
	(0.030)	(0.029)	(0.036)	(0.034)	(0.029)
Number of Claims	0.083**	0.087**	0.096**	0.057**	0.084**
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Length of First Claim	-0.000**	-0.000**	-0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Avg of DV	6.265	6.283	8.114	4.877	6.274
4-digit IPC Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.012	0.012	0.017	0.011	0.012
N	3,689,008	3,882,632	2,782,627	2,424,715	3,877,819

Notes: Unit of analysis is at patent level. Dependent variable is equal to one if the patent is reassigned and zero if not. *Textual Similarity to Science* takes one plus the natural log of the number of times a patent is classified as one of 100 closest patents to a scientific article in Clarivate Web of Science's Science Citation Index-Expanded. The rest of the variable definitions are identical to those in tables 2.6 and 2.7.

In short, our results suggest that the science-MFT relationship is not simply a consequence of unobserved differences in the quality of the patented invention. Our evidence indicates that the science-MFT relationship is stronger for inventions that are based in newer and more novel science, and for more novel inventions. Since transaction costs are likely higher for such inventions, these findings suggest that science based inventions also have higher potential gains from trade. Put differently, the science-MFT relationship is multifaceted, with science potentially reducing transaction costs as well as enhancing gains from trade.

2.6.3 Entry and Market Structure (OLS)

We turn to examining IPC level results. From Equation 2.9 we expect the commercialization capability of the marginal seller (inventor) to decrease as the use of science in an IPC-year increases. Table 2.9 presents the estimation results. We find that IPC-years that have a higher share of patents citing science tend to have a higher share of "small" patentees (column 1), larger number of sellers (column 3), and more first-time patentees (column 5).[51] Our estimates imply that a one standard deviation increase in science-citing patent share from the sample mean translates to a 18% gain in the share of small entities. A similar gain is observed for number of sellers: there are 0.068 sellers per patent on average, but a standard deviation higher citations to science have 0.078 sellers per patent. The share of entrants predicted for IPC-years in the 10th percentile of science-citing patent share is 28%, while it is 32% for patents in the 90th percentile.

[51] Patentees that are patenting for the first time since 1980, divided by number of patents. We exclude the first five years of our panel (1980-1985) to mitigate concern that the early years of the panel will have more entrants.

Table 2.9: Science, MFT, and entry into invention

Dependent Variable:	Small Entity Share		No. of Sellers		Entrant Share	
	(1) Baseline	(2) Novelty	(3) Baseline	(4) Novelty	(5) Baseline	(6) Novelty
Avg(Cite Science Dummy)	0.313**	0.319**	0.060**	0.050**	0.132**	0.119**
	(0.034)	(0.033)	(0.010)	(0.010)	(0.023)	(0.022)
ln(Avg(WOS Combination Familiarity)+1)		-0.003**		-0.001		-0.002*
		(0.001)		(0.000)		(0.001)
Avg(Forward Patent Citations)	-0.010	-0.005	0.021**	0.020**	0.009*	0.009*
	(0.007)	(0.007)	(0.002)	(0.002)	(0.004)	(0.004)
Avg(Triadic Patent Dummy)	-0.370**	-0.333**	0.000	-0.002	-0.079**	-0.072**
	(0.032)	(0.032)	(0.007)	(0.007)	(0.021)	(0.020)
Avg(Number of Claims)	-0.004**	-0.004**	0.001**	0.001**	-0.002*	-0.002*
	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Avg(Length of First Claim)	0.000	0.000	0.000	0.000	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ln(Number of Patents+ 1)	-0.027**	-0.025**	-0.005**	-0.005**	0.007	0.007*
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.003)
Avg of DV	0.281	0.260	0.068	0.260	0.291	0.286
IPCs	334	327	337	328	328	323
Years	31	31	32	32	26	26
R^2	0.928	0.928	0.572	0.591	0.939	0.945
N	6,913	6,290	7,173	6,493	5,899	5,563

Notes: Unit of analysis is at the 4 digit IPC-year level. Observations with fewer than 100 patents are dropped. *Avg Cites to Science* is the count of patents in a 4 digit IPC-year that have made a citations to science. *Small entity share* is the number of small entity (<500 employee) patents. *Number of sellers* equals the number of unique patent sellers that have been identified for each 4 digit IPC-year. *Entrant Share* is calculated by dividing the number of new assignees (entrants) by the total number of assignees in each 4 digit IPC-year. The first five years of the panel are excluded for columns 5 and 6. All columns include fixed effects for 4-digit IPC and patent publication years. Standard errors are clustered at the 4-digit IPC level.

We expect less mature science to be more strongly associated with inventor entry if it affects gains from trade and more mature science if it affects transfer costs. In columns 2, 4 and 6, we include the average of the Scientific Combination Familiarity score calculated at the patent level for each IPC-year. We find that patent classes that are populated by inventions that use more novel (less familiar) scientific combinations are more likely to have smaller entities, more sellers and entrants. This supports the idea that the entrants to scientific invention are capitalizing on gains from trade to sell their inventions.

2.6.4 Instrumental Variable Estimation

Our results that science-based inventions have higher rates of trade, especially for novel inventions, smaller inventors, and heterogeneous markets, is consistent with the view that science lowers knowledge transfer costs and increases gains from trade. This relationship can be confounded by unobserved factors. For instance, an increase in the (unobserved) number of potential buyers would imply an increase in the observed reliance on science. In this section, we obtain causal estimates of the effect of reliance on science on MFT by instrumenting the use of science by changes in U.S. federal contract R&D caused by the end of the Cold War. In the first stage, we predict the share of patents citing science in an IPC-year using changes in federal contract R&D around the end of the Cold War. The predicted values are used in the second stage.

Table 2.10: Post Cold War Federal R&D Shifts and MFT (IV Estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	1st Stage IV	2nd Stage IV	1st Stage IV	2nd Stage IV
Dependent Variable:	ln(Papers)	ln(Share of Reassigned Patents)	ln(Avg Cites to Science)	ln(Share of Reassigned Patents)	ln(Avg Cites to Science)	ln(Share of Reassigned Patents
Δ ln(Gov. R&D Contracting) (WOS Field)	1.257**					
	(0.019)					
ln(Gov. R&D Contracting (Pre, $1Bn)) (WOS Field)	0.364**					
	(0.003)					
ln(Avg Cites to Science)		1.040**		0.905**		0.666**
		(0.056)		(0.145)		(0.105)
Number of Papers (Predicted, 1000s)			0.079**			
			(0.003)			
Δ ln(Gov. R&D Contracting)					0.175**	
					(0.002)	
ln(Gov. R&D Contracting (Pre, $1Bn))		-0.476**	-0.007	-0.478**	0.081**	-0.481**
		(0.086)	(0.028)	(0.106)	(0.016)	(0.081)
ln(Gov. non-R&D Contracting (Pre, $1Bn))		0.314**	-0.059**	0.307**	-0.034**	0.296**
		(0.031)	(0.011)	(0.044)	(0.002)	(0.032)
ln(Number of Patents)		-0.042**	-0.003	-0.037*	0.021**	-0.027**
		(0.007)	(0.004)	(0.015)	(0.004)	(0.007)
Share of Small Assignees		0.396**	-0.245**	0.364**	-0.253**	0.308**
		(0.052)	(0.011)	(0.054)	(0.005)	(0.050)
Avg of DV	5.016	2.059	0.109	2.059	0.109	2.059
SD of Science		0.128		0.128		0.128
Cragg-Donald F-Stat			700.301		360.217	
Year Fixed Effects	Yes	Yes		Yes		Yes
R^2	0.227	0.119				
N	1,373	1,928	1,928	1,928	1,928	1,928

Notes: Analysis for column 1 is at the Web of Science Field-paper publication year level. Analyses for Column 2-6 are at the 4-digit IPC-patent publication year level. Sample period is 1992 and 2000 inclusively for all columns. *Avg Cites to Science* range from zero to one and averages the Cites Science Dummy at the unit of analysis. *Share of Reassigned Patents* ranges from zero to hundred and averages the Reassignment dummy at the unit of analysis, which at the patent level is equal to 100 if a patent is reassigned and zero otherwise. $\Delta log(Gov.\ R\&D\ Contracting)$ is calculated as the difference logged values of R&D contracting obligations between pre (1986-88) and post (1990-92) periods. $Number\ of\ Papers(1000s)$ refer to the number of papers (in thousands) relevant to a 4-digit IPC. $Number\ of\ Papers(Predicted, 1000s)$ is the predicted value from equation A.4. *Gov. R&D Contracting (Pre)* is the average government contract R&D funding for the pre-period (1986-1988). *Gov. non-R&D Contracting (Pre)* is the average government non-R&D contract value for the pre-period (1986-1988). Column 1 includes paper publication year fixed effects, while columns 2-6 include patent grant year fixed effects. Standard errors are clustered at the year level.

Table 2.10 columns 3 and 4 present the results using the predicted paper instrument; columns 5 and 6 present results using the funding differences instrument. Consistent with our prediction of an upward bias, the OLS coefficient (column 2) is larger in magnitude than the second stage IV coefficients in columns 4 and 6. F-statistics for all first stage regressions are above 104.7, which recent work argues to be the appropriate critical F-value for valid inference using a second stage critical t-statistic of 1.96 (Lee et al., 2020).[52] Patent classes that cite science a standard deviation more due to federal research funding shocks experience a 5.6% increase in patent trade probability relative to the sample mean.

2.7 Discussion and Conclusion

This paper aims at advancing our understanding of how science affects the rate and direction of innovation. The use of science in invention can enhance the commercialization of inventions if it facilitates trade in inventions to those that are best able to commercialize them. Such trades support a division of labor between upstream inventors and downstream commercializers. Science generalizes phenomena into universal categories and unravels the mechanisms that underpin them. This may directly lead to higher quality inventions. Furthermore, conceptualizing inventions in scientific terms makes them easier to codify, reducing search costs for potential buyers, and enables buyers to evaluate and integrate inventions. This should reduce transaction costs that are thought to afflict trade in technology, as well as enhance the potential gains from trade.

Our main contribution is to establish that science-based inventions are more likely to be traded. Patents that reference a scientific article are 16-23% more likely to be traded than patents that do not reference science. Trade also depends on

[52]Results are also robust to bootstrapped standard errors with 1,000 samples in appendix table A.5

the demand for inventions. We derive and test three predictions of the reliance on science increasing gains from trade and reducing transaction costs. First, we find that patents invented by smaller firms are more likely to be traded, and that science magnifies this contrast with larger firms. Second, concentrated industries with an unequal distribution of potential patent buyers exhibit more patent trade, especially when the focal patent uses science. Third, the science citation effect on reassignment is up to three times larger for novel patents compared to not-novel patents. We also find that reliance upon science is associated with greater share of small inventors, and with entry of new inventors. Conditional upon citing science, we find that the positive relationship with MFT is especially strong for cited science that is more recent, suggesting that science also increases gains from trade, in addition to reducing transfer costs. Finally, we exploit a variation in the amount of scientific funding available from the federal government in the immediate aftermath of the Cold War to confirm the causal relationship between science and the market for technology.

Our findings imply that enhancing scientific understanding can increase social welfare over and above its role in generating valuable inventions: by encouraging the expansion of markets for technology, which allocates ownership rights to the most efficient user of existing inventions, and indirectly, by supporting a division of innovative labor.

3

The Rise of Scientific Research in Corporate America

3.1 Introduction

Vannevar Bush asserted that science is the "pacemaker of technological progress" (Bush, 1945). Indeed, advances in basic science have served as critical inputs for firms introducing radically novel products such as the transistor.[1] However, in contrast to Bush's vision, private firms often carry out basic research themselves. At the same time, since scientific knowledge easily spills over to rivals (Bloom et al., 2013; Arora et al., 2021b), the reasons why firms may engage in basic science remain an enduring puzzle (Nelson, 1959a). Firms could decide to invest in science to increase absorptive capacity (Cohen and Levinthal, 1989), to attract competent inventors (Stern, 2004; Sauermann and Cohen, 2010), as well as to signal their product quality to buyers (Hicks, 1995; Azoulay, 2002). The present paper explores the role of underdeveloped U.S. universities in the interwar period in inducing firms close to the technological frontier to start investing in scientific research.

[1] *This chapter is adapted from joint work with Ashish Arora, Sharon Belenzon, Konstantin Kosenko, and Yishay Yafeh (Arora et al., 2021a). All authors have contributed equally.*

By the end of World War I, advances in technology were increasingly science-based, especially in industries such as chemicals, plastics, synthetic fibers, electricity and communications. Inventions required more than just trial and error and tinkering; a deeper scientific understanding of the materials, machines and the natural forces at work was needed. General Electric, for instance, had exhausted trial and error methods to reduce the blackening that occurred on the surface of the lightbulb. Irving Langmuir, an American chemist hired by GE after a PhD from the University of Göttingen, made fundamental discoveries in surface chemistry in the course of diagnosing the source of the blackening as evaporation from the tungsten filament under extremely high temperatures. This led to a radically different solution to what had been proposed by technologists at the time: instead of trying to create a better vacuum, Langmuir proposed to fill the bulb with inert gases that would scatter the evaporated particles.[2] Elucidating the science behind existing products also allowed firms to develop valuable new products: DuPont, an explosives manufacturer that had by the 1920s diversified into paints, rubbers and rayon, all of which were based on polymeric materials, invested in polymer science, which yielded blockbuster products such as nylon and polyester.

Around the same time period, in the early decades of the twentieth century, several large American firms had reached the technological frontier in their respective markets and realized the need for a more fundamental understanding of how their flagship products worked. However, as we show below, the scientific knowledge available from American universities was inadequate for firms at the technological frontier because of the relative backwardness (in comparison with Europe) of U.S. universities, especially in certain scientific disciplines within physics and chemistry.[3]

[2]Langmuir recalled that his work on surface chemistry allowed him to "conclude with certainty that the life of the lamp would not be appreciably improved even if we could produce a perfect vacuum" (Reich, 1983).

[3]Indeed, American oil companies had sponsored the establishment of chemical engineering as a

We argue below that the firms that felt this need most acutely were large, publicly traded firms, which relied on novel inventions for their growth. There is also some evidence to suggest that they operated in relatively concentrated markets, where growth through mergers and acquisitions was increasingly circumscribed by antitrust laws.[4] Continued sources of growth had to be found in new products and markets. Because the public science system was underdeveloped, these firms chose to invest internally. Because they operated in concentrated markets, spillover of knowledge to rivals was apparently not a major concern.

The relationship between the state of academic research and corporate investment in science is complex. Academic science could complement or substitute for internally generated research by companies. Moreover, because academic science is potentially available to all firms in a market, the nature of the strategic interactions among competitors matters as well. Finally, human capital is jointly produced along with academic research. The combined effect is complicated and likely to be historically contingent. We develop a simple conceptual framework to study the private returns to investment in research. We distinguish between scientific knowledge and innovation. Innovation–the introduction of new products and processes–is the source of profits. Scientific knowledge, both from universities and from internal research, reduces the cost of innovation. Leading firms that are more dependent on innovation derive greater returns from investing in research. However, their incentives also depend upon the nature of strategic interactions, as well as on the state of academic science. Under some conditions, the incentives to invest may actually be higher when the supply of academic science is low.[5]

discipline at MIT because they needed to increase refining efficiency.

[4]DuPont had grown by acquiring smaller innovative firms such as Hercules, and Giant, which had the license from Nobel. Following the breakup of the rail and oil trusts, DuPont had to divest its existing operations in smokeless powder in 1911.

[5]The model does not consider the effect of the production of human capital. See Arora et al.

We study these issues empirically using newly assembled historical data, which combine novel data on corporate innovation with ownership and financial information during the interwar period. We assemble three measures of innovation: publications, industrial lab employment and patents. We collect data on scientific publications authored by researchers employed at corporations using Microsoft Academic Graph (MAG), including the scientific discipline they publish in, as well as citations received by other publications and patents. We match firms to the *Industrial Research Laboratories of the United States* survey (hereafter "the IRL directory") of corporate science laboratories. A total of 306 firms (65%) in our sample of patenting firms operate a research lab, while 196 (42%) publish at least once during our sample period. We complement the information on corporate science by matching our sample firms to U.S. patents[6] and link them to in-text citations made to scientific publications published during this period (Marx and Fuegi, 2020).[7] Our dataset includes not only the number of patents by company but also measures of patent quality, such as forward patent citations received and stock market values from Kogan et al. (2017).

To measure the public science that firms could draw upon, we develop measures of relative American backwardness by scientific field, as well as measures of a firm's proximity to the technological frontier. Scientific backwardness is measured using three alternative methods. In the featured method, we calculate for each scientific field the number of academic publications by European authors in the field, divided by all publications by U.S.-based scientists in the same field.[8] Scientific fields are

(2019a) who analyze how the joint production of knowledge and human capital conditions the incentive of a single incumbent in a model where the incumbent may potentially buy inventions from startups.

[6]Provided through Google Patents and IFI claims services.

[7]Front-page NPL citations were introduced in 1947 at the U.S. Patent Office.

[8]In the second measure, we calculate the share of citations (out of total backward citations) made to European journals by American journals in each scientific field. For the third method, we calculate the ratio of prominent American scientists trained in Europe divided by American scientists without experience in Europe.

weighted by their relevance to the firm's patenting activity. A firm's proximity to the technological frontier is measured through prior-art citations a firm's patents receive, in-text citations its patents make to scientific articles and the ratio of patents to assets. The combined data set comprises 469 firms and 7,035 firm-year observations.

We combine the data on corporate innovation with firm-level panel data, which include detailed financial information and ownership. We extend a dataset constructed by Kandel et al. (2019) (hereafter "KKMY"), which includes both publicly traded and private firms that were active in upstream R&D (academic publications) or downstream R&D (patenting) during the period 1926 to 1940.[9] We use information on corporate ownership and control. In particular, we identify firms that are controlled by other (parent) firms, or control their own subsidiaries, and firms which are part of business groups with multiple affiliates, often operating across industries. Such interfirm ties matter because, prior to 1940, U.S. corporate ownership was quite different from what it is today; business groups and conglomerates of various types dominated the U.S. economy.[10] We supplement the information on corporate ownership and organization with data on firm size (assets) and profitability, collected from Moody's Manuals and from the Center for Research in Securities Prices (CRSP).[11] We also include data on the extent of competition in each firm's 3-digit industry using a cross-sectional classification of sectors into monopolistic, oligopolistic and competitive ones (Wilcox, 1940).

Our sample period, ranging from 1926 to 1940, is well suited to the issue at

[9]Companies were included only if they had at least one publication or one patent over the sample period.

[10]For instance, the DuPont company, a corporate pioneer in chemical research, also had a significant ownership stake in General Motors. Indeed, Pierre DuPont became President of GM in 1920 after having served as DuPont's President. DuPont invented and produced lacquers and enamels for the growing automotive sector. It also manufactured tetraethyl lead, the anti-knock agent discovered by Thomas Midgely, a chemist working at GM.

[11]Moody's Manuals provide balance sheet data, including sales and assets. CRSP contains information on a subset of listed firms, for which market capitalization is calculated.

hand.[12] First and foremost, this is the period in which the phenomenon of corporate science emerged in the United States. Second, our sample period ends before the onset of World War II, when the U.S. government became deeply involved in scientific development both directly and through military procurement contracts (Gross and Sampat, 2020a,b). During our sample period (much of which coincides with the Great Depression), the U.S. government had little influence on corporate research and development activities.

We find that firms in the technological frontier were more likely to invest in science. For instance, firms with forward patent citations above the sample average published around nine times more in scientific journals than firms whose patent citations were below average. The difference is similar for industrial lab sizes: firms with above-average forward patent citations employed around ninety lab personnel on average, while firms with below-average citations employed eleven. Our results also indicate that large firms, operating in concentrated markets, and affiliated with business groups published more. Moreover, these relationships are stronger for firms that relied on scientific areas in which American public science was relatively less developed. For instance, firms whose patents cited science themselves published more, especially when the firms relied upon scientific fields, such as chemistry and physics, that were relatively underdeveloped in the United States.

We make three contributions to the literature. First, we provide empirical evidence on, and explanations for, the rise of corporate research in America and the emergence of an innovation ecosystem, which played an important role in promoting American science and the U.S. economy to world prominence. Lamoreaux and Sokoloff (1997) study the internalization of R&D by U.S. firms around the turn of the twentieth century from the perspective of the independent inventors and their

[12]Our sample begins in 1926, as this is the first year in the KKMY sample and the year in which the CRSP data on stock market data becomes available.

gradual conversion into salaried R&D employees in larger firms. Nicholas (2010, 2014) attributes one of the causes of this transition to the increasing complexity of the technology underlying the chemical and electric industries. Company histories (Reich, 1985; Maclaurin and Harman, 1949; Jenkins and Chandler, 1975; Hounshell and Smith, 1988; Hounshell, 1990, 1998) document the various motives behind the establishment of large industrial R&D laboratories. Mowery (2009) notes that companies were withdrawing from research by the 1980s, and Arora et al. (2019a) show that this trend reflected changes in firm behavior and was not simply due to changes in size or composition of industrial activity. We show that the rise of corporate research was spearheaded by firms operating on the technological frontier, typically large, publicly traded firms, some of them affiliated with business groups, especially in scientific fields that were underdeveloped in the United States. This is consistent with the interpretation that the weakness of university research provided an opportunity for firms to gain competitive advantage by internalizing scientific research. Our findings suggest that the dramatic growth of university research after WWII affected the private returns to firms from undertaking such research themselves.

Second, we provide evidence in support of the missing institutions (or institutional voids) framework in a context hitherto not documented in the literature. Khanna (2000, 2018), and Khanna and Palepu (2000) apply this idea to emerging markets; the present paper finds that this concept may be useful in understanding phenomena related to the U.S. economy in the first half of the twentieth century. Our results are consistent with the view that corporate research in America was driven by the attempts of large corporate entities to make up for the weakness of American academia at the time. This finding, as well as the finding that business group affiliates were prone to invest in basic scientific knowledge, are consistent with Nelson (1959b). In addition, these results provide a historical perspective on contemporary debates regarding the role of large U.S. corporations in advancing science and promoting

innovation. Whereas Autor et al. (2020) describe large U.S. corporations as highly innovative "superstar firms," others, such as Gutiérrez and Philippon (2020), view them as primarily inefficient entities shielded from competitive pressures.

Third, the dataset we construct, combining information on U.S. corporate ownership and U.S. corporate science in the interwar period, is the most comprehensive of its kind. This dataset should open the way to future research on the possible links between corporate characteristics, research and development, government policy and institutional context.

The paper is organized as follows. In the next section, we summarize the rise of American corporate science in the early twentieth century. Section 2 and 3, respectively, survey the historical context and set up the theoretical framework to for our analyses. Section 4 describes our data. Section 5 presents our econometric specifications and estimation results. Section 6 concludes.

3.2 The Rise of Corporate Science in America

Prior work has established that American firms in the early twentieth century steadily increased the scope of their activities, and some firms invested in internal R&D as well (Chandler Jr, 1977). Lamoreaux and Sokoloff (1997), in their study of American inventors between 1870 and 1911, show that independent inventors who had previously supplied inventions and contract research to firms began to be directly employed by firms. Between 1921 and 1940, the number of firms with labs increased more than sevenfold, from 297 to 2264 (Mowery and Rosenberg, 1998). The available evidence from company histories suggests that in the early years, corporate labs focused on quality control and solving operational problems rather than fundamental science.[13] For instance, Charles Dudley's tenure at the Pennsylvania Railroad com-

[13]Robert Duncan, the founder of the Mellon Institute for Industrial Research (established in 1913, the premier contract R&D organization of its time) and an advocate of corporate research,

pany that began in 1875 was focused on examining the metallurgical properties of the rail tracks that were supplied to the firm by steel companies. Thomas Edison's Menlo Park facility was known for shunning fundamental investigations that Edison considered "purely aesthetic" (Wise, 1985).[14] However, by WWI, inventions were increasingly science-based. Initially, companies looked, as they had in the past, to external suppliers to fulfill this need.

This motivated the establishment of specialized contract research organizations, such as the Mellon Institute in 1913. The institute grew steadily in contract revenues ($300,000 to $800,000) between 1919 and 1929. Over the same period, the number of industrial fellows sponsored by firms grew from 83 to 145. Industrial fellows, such as George Curme, made crucial contributions to replacing coal tar with petroleum for certain fine chemicals, while Union Carbide's contract with the institute yielded ethylene gylcol, the antifreeze, which became a key product for the firm (Servos, 1994, 223). But contract research worked best for generic, well-specified problems. Outsourcing research entailed contracting problems and required the costly transfer of firm-specific information. Indeed, outsourcing research required that the firm itself had significant research capabilities (Mowery, 1983).

Research managers, such as Willis Whitney at GE Research Labs, Frank Jewett at AT&T Bell Labs, CEK Meese at Kodak and Charles Stine at DuPont, therefore chose to invest internally. Other firms, such as Westinghouse, Standard Oil, Western Union, RCA and Alcoa, soon followed by also instituting in-house research programs. As expected, the quantity of corporate research increased. As we show in Section 3.4, American firms published three times as many scientific papers in 1940 as they did

lamented that the factories of American firms were dominated by foremen that stuck to traditional practices and that managers were too myopic to wait the "two, three, or even five years" for scientific projects to reach their potential (Servos, 1994, 223-225).

[14]According to a foreign-trained lab employee, Edison argued that, "We can't be like those old German professors who, as long as they can get their black bread and beer, are content to spend their whole lives studying the fuzz on a bee." (Wise, 1985)

in 1926. The quality of research also improved dramatically: Irving Langmuir (GE) was awarded his Nobel Prize in Chemistry (1932) and Clinton Davisson (AT&T) in Physics (1937).

3.2.1 Causes Behind Corporate Science

Why did firms invest in science? There are four explanations, which are not mutually exclusive and perhaps even complementary.[15] First, German chemical firms such as Bayer, Hoechst, and Agfa fared well in the international organic synthetic dyes market by building on their corporate research (Smith and Hounshell, 1985). This set a precedent for American firms to emulate (Reich, 1985, p.41).

Second, American inventions were being challenged by European competition, which was leveraging scientific advances to invent new products and processes. GE's electric lighting business that was started by Edison in 1879, for instance, was based on carbon-filament high-vacuum incandescents. However, Walther Nernst (the 1920 Nobel Laureate in Chemistry) at Göttingen invented a glower that required no vacuum to operate and was more efficient. Westinghouse eventually acquired the patent rights to the Nernst glower in a bid to compete with GE (Wise, 1985). This was one of the motivations behind the establishment of GE Research Laboratory (GERL) in 1900. Langmuir, one of Nernst's American doctoral students, was also recruited by Whitney to further his work on surface chemistry.

Third, American firms had often reached the technological "frontier," and improving their existing products and processes required a deeper scientific understanding of how they worked. New products in a range of industries, such as chemicals, electric lighting and communications, were even more deeply rooted in scientific advances, but few independent inventors had the required scientific capability and equipment. When firms attempted to acquire technology from abroad, a lack of scientific sophis-

[15]This section draws on Arora et al. (2020).

tication prevented them from fully exploiting such imports. For instance, after the U.S. entry into World War I, the federal government mandated compulsory licensing of German patents under the Trading with the Enemies Act (TWEA) of 1917. However, American firms found it difficult to replicate the products described in the patents. DuPont and other U.S. firms had to expend substantial R&D dollars to reproduce German dyestuffs, underscoring the need for internal research (Bhardwaj et al., 2006; Hounshell and Smith, 1988).[16]

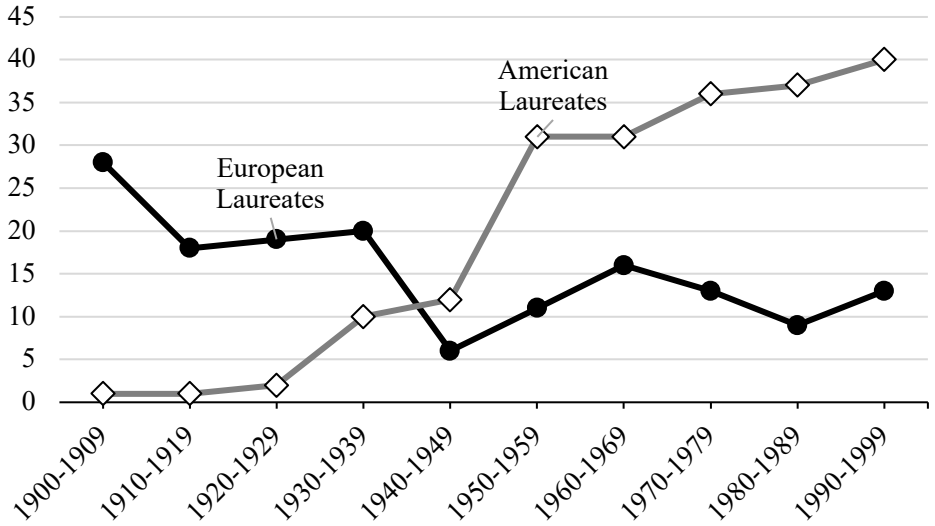


FIGURE 3.1: Number of Natural Science Nobel Prize Laureates, by Citizenship at Award[17]

Finally, in contrast to the European system, American universities were less focused on basic research and lagged their European counterparts, particularly in quantum physics and organic chemistry in the interwar period.[18] Five out of the 42 Nobel

[16]Inadequate disclosure in patents did not help. A classic example was the Haber-Bosch process, which was critical for synthesizing nitrogen, where BASF had withheld information about the catalysts required (Haynes, 1945).

[17]The line graph plots the number of total laureates in the Nobel Prize for Physics, Chemistry, and Physiology/Medicine. The home countries of the winners are coded based on the classification by the Encyclopedia Britannica (please see https://www.britannica.com/topic/Winners-of-the-Nobel-Prize-for-Physics-1856942 for page for Physics). According to the source, "Nationality given is the citizenship of recipient at the time award was made."

[18]Irving Langmuir, for instance, was disenchanted with the lack of research support at Stevens

Prizes in Physiology (12%), three out of the 39 in Chemistry (8%) and six out of the 46 in Physics (13%) awarded between 1901 (the first awards) and 1939 went to American scientists. This is in stark contrast to the post-1940 period (1940-2020), where the American share jumps to 55% for Physiology, 68% for Chemistry, and 51% for Physics (see Figure 3.1 for trends over time). The gap was wider in select fields of chemistry and physics: only two Americans (Irving Langmuir and Karl Compton) were invited to the famous 1927 (5th) Solvay Conference on Electrons and Photons. In mathematics, Germany produced well over half of all PhDs in the subject until 1920, while America was responsible for around a quarter (Castelvecchi, 2016).[19] This is indicative of the types of transatlantic intellectual interactions that occurred before the 1930s: American physicists would receive postdoctoral training in Germany, while German physicists would come on lecture tours in the United States (Holton, 1981; Fleming and Bailyn, 1969).[20]

University research budgets also reflect this gap. U.S. universities before WWII received very little federal funding for research. Geiger (1986, pp. 273-4) notes that the number of publications by the nation's top 16 universities at the end of World War I was less than a quarter of its level at the eve of World War II. In 1919, the budgets of these universities were less than half of their level in 1937. Firms may have

Institute of Technology, where he joined as a faculty member in 1906: "To his chagrin, he found few students with an interest in science. His attempts to upgrade laboratory facilities and the quality of student work met with hostility from students and indifference from his colleagues. To make matters worse, he had little time for research. When he left in 1909 to join the GE Research Laboratory, Langmuir found a position that met his needs far better" (Reich, 1985, p.111).

[19]The U.S. would overtake Germany as the largest producer of mathematics PhDs by 1940.

[20]To be clear, there were areas of American scientific and technological excellence. Agricultural sciences and mechanical and civil engineering, for instance, were nurtured after the Morrill Act of 1862. There was a long tradition of applied research for specialized products, such as boilers and rubber in Purdue University and the University of Akron, respectively (Geiger, 1986). In machine making, American firms were technologically advanced to the extent that German competitors were playing catch-up by reverse engineering imported American machines (Richter and Streb, 2011). Moreover, American universities were clearly catching up with European universities during the interwar period (Urquiola, 2020; MacLeod and Urquiola, 2021).

therefore found research from American universities inadequate for their purpose.

A comparison of the research expenditures between the best corporate laboratories and universities of the era underlines this point. In its "Research: A National Resource" report published in 1938, the National Resources Planning Board under the NRC surveyed 1,450 American colleges and universities and found that the top 150 spent an average of $ 333,333 per university on research (Council, 1938). The University of Chicago ($2,557,803 in 1929-30), and the University of California ($2,350,000 in 1928-29) were the top research spenders. By comparison, Hounshell and Smith (1988, p. 612) note that DuPont's 1925, 1930 and 1935 budgets were $1.99 million, $5.5 million and $6.6 million, respectively. This implies that DuPont's R&D budget in 1930 was as large as that of the Universities of Chicago and California put together. AT&T's R&D data from Maclaurin and Harman (1949, p. 158) show that the 1925, 1930 and 1935 budgets were $11.7 million, $23.2 million and $15.4 million, respectively. Simply put, the weakness of American university research represented a gap, so companies such as Standard Oil, DuPont, AT&T, GE, Kodak and Alcoa could hope to gain sustained advantage by creating scientific expertise in-house.

3.3 Theoretical Framework

The foregoing account of the rise of corporate research stresses three factors: the imperative to innovate for the leading firms, the role of science in facilitating innovation and the weakness of American university science. To study more formally how these factors interact, we adapt the framework developed in Arora et al. (2021b). Whereas they analyze the impact of spillovers, we focus on the differences across firms in the payoffs from innovation and the effect of public science on research investments.

There are two firms, indexed by 0 and 1. Both compete in the product market, and both invest in innovation, d_0 and d_1, respectively. There are three stages. In stage 3, the firms compete in the product market. Their product market performance

depends on the quality of their products and the cost of producing them. We assume that cost and quality depend upon the innovation output, d_i, $i = 0, 1$. Their payoffs from stage 3 are $\Pi(d_0, d_1)$ and $\tilde{\Pi}(d_1, d_0)$, where the tilde indicates firm 1. Firm 0 is closer to the frontier, so that its marginal return from innovation is higher than that of its rival. Firms farther from the frontier can increase profits by imitation, reducing production bottlenecks and increasing scale, possibilities that the leaders have already exhausted. Instead, leaders have to introduce new and improved products and processes—to innovate. Accordingly, the marginal product of innovation for firm 0 is greater than that of firm 1. To represent this, we assume that the payoff from innovation for firm 0 has an additional term, $kd_0,\ k > 1$.

In stage 2, firms choose their innovation output. Firm 0 chooses d_0 and firm 1 chooses d_1. The cost of innovation for firm 0 is $\phi(r_0; u)d_0$, where r_0 represents investments in internal scientific research by the firm and u indexes the stock of (relevant) public science. Innovation typically requires the invention of new products and processes. Internal research reduces the cost of invention by guiding the search for inventions in more promising directions. Innovations may also be based on inventions acquired from independent inventors, other firms or university researchers. Thus, the cost of innovation also depends on the state of public science. It is natural to assume that both internal research and public science reduce the unit cost of innovation, $\phi(r_0; u)$, i.e., $\frac{\partial \phi}{\partial r_0} < 0$, $\frac{\partial \phi}{\partial u} < 0$, and diminishing returns so that $\frac{\partial^2 \phi}{\partial r_0^2} > 0$.

In stage 1, firms may invest in research. Firm 0's research investment is denoted by r_0, and the cost of research is modelled simply as $\frac{\gamma}{2} r_0^2$, so that the value of the firm, $v = d_0 - \frac{c_{00}}{2} d_0^2 - b d_1 - \frac{c_{11}}{2} d_1^2 + c_{01} d_1 d_0 - \phi(r_0, \lambda) d_0 - \frac{\gamma}{2} r_0^2$.

3.3.1 Model Predictions and Empirical Implications

We provide details and proofs in Appendix C.1. Here we provide the main results and the intuition. Although the profit and cost functions are otherwise symmetric, this

additional term for the leader can result in markedly different outcomes. The returns to investing in research depend on the scale of innovation because research reduces the unit cost of innovation. Firm 0 has a higher marginal return to innovation, in equilibrium, it will innovate more, and thus will have a higher marginal return from research. Furthermore, as k, representing the higher returns from innovation to the leader, increases, the marginal return to investing in research increases for the leader. The supply of public science will enhance the effect of k on research if it complements research and will diminish it if the two are substitutes in reducing the cost of innovation.

As k increases, the returns to research fall for the follower if innovations are strategic substitutes. Intuitively, an increase in innovation by the leader reduces innovation by the follower if innovations are strategic substitutes. The decline in its innovation reduces the follower's incentives to invest in research. In other words, the gap between the leader and follower in the value of innovation leads to a corresponding divergence in the marginal returns to investing in research if innovations are strategic substitutes.

The first empirical implication of the model is that there is likely to be an important extensive margin for research. Indeed, as we show in Section 3.4, the distribution of innovations across firms is highly skewed, and, further, only a minority of firms that innovate invest in research. We also find that firms with higher returns from innovation–e.g., more patent intensive firms and firms with higher quality patents–are more likely to invest in research.

Related to this, as the gap between the leader and follower increases, the leader's returns to investing in research will depend on the supply of public science, increasing with public science if public science and research are complements and decreasing with public science otherwise.

When we restrict attention to the intensive margin, research investment depends

on two other factors: public science and strategic interactions in the product market. An increased supply of public science would tend to increase investment in research, unless public science is a substitute for internal research. Strategic interactions are important as well. If innovations are strategic substitutes, so that an innovation by the rival reduces the marginal returns to innovation by the focal firm, then public science would increase internal research even if public science and internal research were independent in their direct effect on innovation. Put differently, if we observe a negative relationship between public science and internal research, it implies that public science and research are substitutes, or that innovations are strategic complements, or both. Finally, public science can reduce the value of the leader, especially if public science and internal research are substitutes in reducing the cost of innovation.

3.4 Data

Our unbalanced panel of firms is constructed by matching several datasets: the corporate ownership and financial statements dataset assembled by KKMY,[21] augmented by market value data on other listed companies from CRSP; USPTO data from Google Patents and publication data from Microsoft Academic Graph (MAG). The combined dataset covers the period 1926-1940.

We begin with 234 firms from KKMY that patent at least once within our sample period in an IPC that cites at least five scientific articles between 1947 and 1957. This restricts our sample to firms that are "at risk" of beginning scientific research.[22] We augment KKMY, which consists of large industrial firms active during the 1920s, by including 235 listed firms from CRSP that patent (Kogan et al., 2017). Therefore,

[21] As KKMY collect data for 5 years (1926,29,32,37,40), we collect data for the intervening years through the Moody's Manuals.

[22] Examples of excluded patent classes include B27M (woodworking), B60P (loading transportation vehicles) and E03D (Water Closets or Flushing Valves thereof). Around 26% of patenting firms in our sample are lost to this restriction.

our basic sample consists of 469 private and public American firms (7,035 firm-years) that patent at least once in our sample period in an IPC that cites at least five scientific articles between 1947 and 1957. Of these, there are 469 firms (and 4,305 firm-years) for which we have financial statement data.[23]

3.4.1 Scientific Publications

We use Microsoft Academic Graph (MAG) to source 283,992 peer-reviewed scientific publications between 1926 and 1940. We exclude papers in the humanities and the social sciences based on their OECD subfields.[24] We calculate publication stock for a firm-year using a perpetual inventory method with a 15% depreciation rate. Similar to patents, we calculate normalized forward publication citations by dividing raw forward citations received by publications up until 2019 by the average number of forward citations received by the focal cohort (papers published in the same year). Using Marx and Fuegi (2020), we also identify which publications are cited by a U.S. patent in the future.[25]

We distinguish between university authors and corporate authors of scientific publications. We match 140,766 author affiliations from 283,992 papers to each sector. For universities, we filter publications that contain affiliations that indicate academic authorship, such as "University," "College," "Institute of Technology," "School" etc. This process yields 60,305 affiliations (123,657 papers) that are related to universities for 1926-1940. We limit our sample to affiliations in the United States only. For corporations, we use a fuzzy string matching algorithm that takes into account abbreviations frequent in the era (e.g., firms in the railroad sector may

[23]The difference in firm-years is because patent data are available for all years in our sample period, but financial data have years without coverage.

[24]These fields of science (FOS) have been defined in the 2002 revision (6th edition) of the Frascati Manual. See https://www.oecd.org/science/inno/38235147.pdf for full list of classifications.

[25]We use Version 24 of this dataset, available from https://zenodo.org/record/3976926#.YSFi2S2cZTY

be abbreviated as RR (railroad), RW (railway), RC (rail company)), and name variants for certain companies (e.g., AT&T's Bell Labs, SOCONY for the Standard Oil Company of New York). We ensure that eponymous charitable foundations and hospitals (e.g., by DuPont, Carnegie and Rockefeller) are not erroneously classified as corporate publications. We match 3,194 corporate publications to 201 sample firms. Of these, 110 are found in KKMY sample, 162 are found in CRSP and 71 are found in both samples.[26]

In Table 3.1, we find that electrical engineering was the scientific field in which the most papers were published, with a total of 1,642 papers published by 156 firms during our sample period. The top publishers in Figure 3.2, GE, AT&T and Westinghouse, all publish heavily in this field. Electrical engineering is followed by physics (461 articles) and chemistry (344 articles). The field with the higher number of forward citations received from other papers is physics (1.64), followed by biology (1.37).

[26]See Appendix B.3.5 for details on matching.

Table 3.1: Corporate Scientific Publications, by OECD Subfield

OECD Subfield	Number of Firms	Number of Papers	Avg Forward Publication Cites
1.03 Physical sciences and astronomy	63	461	1.64
1.06 Biological sciences	32	83	1.37
2.03 Mechanical engineering	89	186	1.12
2.02 Electrical eng, electronic eng	156	1642	0.75
1.04 Chemical sciences	76	344	0.72
1.07 Other natural sciences	2	2	0.60
1.02 Computer and information sciences	36	85	0.59
2.05 Materials engineering	56	186	0.54
1.01 Mathematics	36	95	0.52
3.02 Clinical medicine	50	139	0.47
4.01 Agriculture, forestry, fisheries	13	19	0.45
1.05 Earth and related environmental sciences	64	136	0.35
2.08 Environmental biotechnology	9	9	0.34
2.11 Other engineering and technologies	61	119	0.33
2.06 Medical engineering	8	25	0.27
3.01 Basic medical research	23	30	0.22
2.07 Environmental engineering	96	259	0.20
2.04 Chemical engineering	21	24	0.18
3.03 Health sciences	17	27	0.17
4.02 Animal and dairy science	8	11	0.14
2.01 Civil engineering	57	106	0.14
4.03 Veterinary science	2	3	0.04
4.05 Other agricultural science	12	14	0.02
Not Available	62	131	0.01

Notes: Observations are at OECD subfield level for years between 1926 and 1940. "Number of Firms" counts the number of firms publishing at least one article in the focal field. "Number of Papers" counts the number of total papers in the focal field. "Average Forward Publication Cites" take the field-level average of the normalized forward citations. Forward citations are normalized by the average number of forward citations received by all publications published in the focal publication's year.

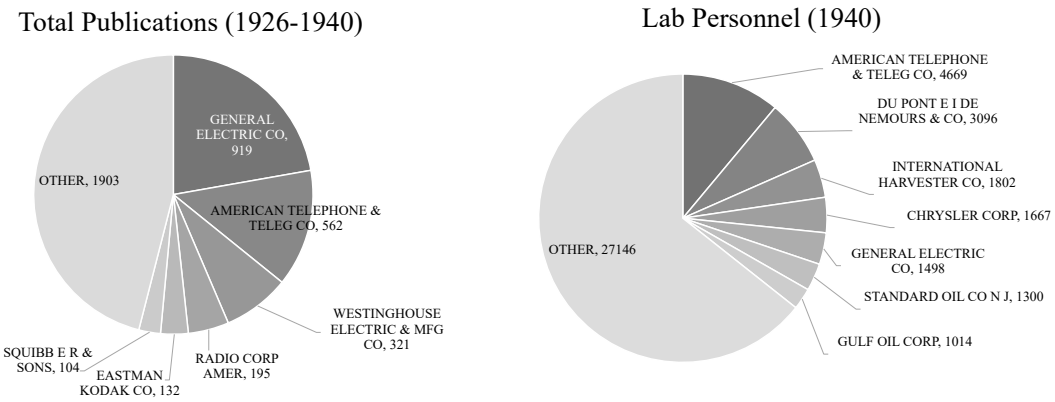


FIGURE 3.2: Firm Share of Publications and Personnel[28]

3.4.2 Corporate Labs

We also obtain data on the size of R&D labs operated by firms from a national survey by the National Research Council (NRC) conducted since 1920 (Service, 1931). Data from these surveys have been used in Mowery and Rosenberg (1998), Nicholas (2011), Field (2003) and Furman and MacGarvie (2007). We manually match our firm to firms in the 1927 (999 firms), 1931 (1,620 firms), 1933 (1,562 firms) and 1938 (1,769 firms) surveys. We collect the number of total personnel that were employed in labs. Figure 3.2, right panel, shows the distribution of lab personnel by firm. Some of the top publishers (shown in the left panel), such as AT&T and GE, also operate the largest labs, but the rank correspondence is not one-to-one. For instance, DuPont operates a large lab with over 3,096 personnel, but publishes a total of 21 papers in our sample period. This may reflect a heterogeneity in firm publishing policies, as well as field-specific publishing behavior.[27] This underlines the importance of normalizing the number of American scientific papers (from WoS) by European counterparts to account for field-specific differences in publication behaviors.

[27]For instance, Western Union's 1931 publication in the Transactions of the American Institute of Electrical Engineers reports the construction of a new transatlantic cable the firm laid in 1928 (See https://ieeexplore.ieee.org/document/5055804).

3.4.3 Gap in University Science

We measure the "void" or "gap" in university science in America compared to Europe by the citation-weighted scientific publications authored by scientists in each region. We also find broadly similar results using two alternative measures: the number of scientists trained at or affiliated with a European university, and the citations to European journals made by American journals.

Scientific Publications: U.S. and Europe — We use the country of correspondence for the authors of scientific publications. We first collect address information of authors for 44,355 publications published between 1900 and 1920 from WoS and classify addresses into US, Europe and "Rest of World" regions based on their country names.[29] For publications missing addresses, we match the authors' last and first names to the *American Men of Science* directory to identify 27,924 publications by prominent American scientists. The rest of the publications during this period are classified as European. We exclude papers in the social sciences and humanities and are left with 15 OECD subfields for which at least one "European" or "American" published between 1900 and 1920.[30] The above process yields 155,571 publications by Europeans and 60,605 publications by Americans in the sciences between 1900 and 1920. To adjust for quality differences, we weigh the publication counts by the number of forward paper citations received until 2019. These numbers are broken down by field in Appendix Table B.17.

[28]The left pie chart downward sorts the number of total publications by firms in our sample clockwise. The right pie chart sorts the number of lab personnel reported by firms in the 1940 edition of the Industrial Research Laboratory directory clockwise.

[29]Microsoft Academic Graph does not contain an address field, whereas Web of Science contains a separate field dedicated for addresses and country classifications based on these. See Appendix Tables B.8 and B.9 for details on classifications.

[30]We use the correspondence in Marx and Fuegi (2020) to map Web of Science subject fields to 39 OECD subfields.

American Men of Science — Scientific areas where the United States is ahead will exhibit more homegrown scientific talent, while areas where the U.S. lags behind Europe will feature more scientists trained in Europe. For instance, the founder of the *American Journal of Chemistry* (Ira Remsen) studied at Göttingen, while the alma mater of the founder of the *American Journal of Mathematics* (James Sylvester) was the University of Cambridge (Kevles, 1979). On the other hand, areas such as agriculture and civil engineering, where the United States did not lag as far behind, did not require a similar import of overseas talent. We collect information on European education/affiliation by American scientists from the 1921 (3rd) edition of the Cattel Directory of American Men of Science.

Published by James McKeen Cattel since 1906 and running its 38th edition in 2020, the American Men of Science directory (hereafter AMS) is one of the oldest and most comprehensive listings of scientists active in the United States (Moser and San, 2020). To measure relative scientific strength before our sample period (which starts from 1926), we focus on the 1921 edition because the Optical Character Recognition (OCR) quality is highest, and it provides the most comprehensive listing of scientists.[31] The number of listed scientists increases from around 4,000 in 1906 to 5,500 in 1910 and 9,500 in 1921, likely reflecting both the growth of American science as well as better coverage by Cattel. We extract 8,232 author entries. Of the 7,245 scientists on whom we have the required information, 1,649 are trained in European institutions.[32] It is possible that scientists who were exclusively trained in the United States up to the doctoral level are recruited by European institutions and show up as "European" due to their affiliations, though random checks suggest that this is quite rare. We manually classify each scientist into a scientific field (OECD subfield) based on the subject listed for each entry (Appendix Figure B.6 shows an

[31] https://catalog.hathitrust.org/Record/003255132 for details.

[32] See Appendix B.4.1 for details on cleaning the AMS data.

entry for a scientist listed under chemistry). This allows us to count the number of European-affiliated (and non-affiliated) scientists in America.

Web of Science: Transatlantic Journal Citations — Another way to measure scientific gaps is by the number of citations made to publications in European journals by American journals. We classify 244 journals indexed in the WoS Science Citation Index - Expanded (SCI-EXPANDED) as "American" or "European" based on name and web searches. Journal names in non-English languages, such as "Comptes Rendus" and "Zeithschrift für Physik," are first identified as non-American (with the exception of those in Latin such as "Acta Mathematica"). All other journals are searched online and classified based on the home country of the academic society. Where this information is not available, we use the home country of the journal's founders. 230 journals out of the 244 are classified, 111 (45%) of which are American.

For articles published between 1900 and 1920, we count the number of citations made by "American" journals to "European" journals in the same period. This constitutes a measure of European scientific strength — if a field relies more on European science, citations to European journals would be higher. We also count the number of citations made to American journals, which constitutes a measure of American scientific strength.

Calculating Scientific Gaps for Firms — Our regional scientific activity data (from AMS and WoS) are encoded at the scientific field level. Therefore, we link them to firms based on how much each firm patents in a patent class, and on how much a patent class relies on a scientific field. For instance, for the first measure using publication addresses, we calculate the number of papers (European and American) published between 1900 and 1920 relevant for each 4-digit IPC based on the share of patents in that IPC that cite the OECD subfield in their front page Non-Patent

Literature section:[33]

$$European\ Papers_{IPC,field} := \frac{NPL\ Citations_{IPC,field}}{NPL\ Citations_{IPC}} \times European\ Papers_{field} \tag{3.1}$$

We sum $European\ Papers_{IPC,field}$ over all OECD subfields to obtain the number of (European) papers "relevant" to a given IPC: $European\ Papers_{IPC}$. We then map this IPC level value to a firm using the share of patents the firm has in each IPC over the sample period, 1926-40.

$$European\ Papers_{firm,IPC} := \frac{Patents_{firm,IPC}}{Patents_{firm}} \times European\ Papers_{IPC} \tag{3.2}$$

Summing $European\ Papers_{firm,IPC}$ over the 4-digit IPCs, we obtain the number of European publications relevant to each firm. We repeat the same procedure for American publications published between 1900 and 1920 in WoS. We then divide the number of European publications by the sum of the American and European publications at the firm level to get our primary measure of scientific gap the firm faces.[34]

For the gap measures using scientist affiliations in the 1921 edition of the American Men of Science, we replace $European\ Papers_{field}$ in Equation 1 with the number of scientists in that OECD subfield that have been trained, at least in part, in Europe (while $American\ Papers_{field}$ are replaced with the scientists without European

[33]We use data for patents granted for the first 10-year period from the time NPL citations were formalized in U.S. patent documents (i.e., between 1947 and 1957).

[34]For example, AT&T, which is in the 90th percentile in this score, 15% of its patents in IPC H01J (Electric discharge tubes or discharge lamps) between 1926 and 1940. Patents in this IPC, in turn, cite the Chemical Sciences most often (26%), followed by Electrical Engineering (23%) and Physical Sciences (21%) between 1947 and 1957. As we see in Table B.17, Chemical Sciences and Physical Sciences have European-to-American ratios that are higher than the average, which contributes to the higher firm-level gap score for AT&T. In contrast, General Ice Cream Corp, which is below the 10th percentile in this score, patents most often in A23G (Cocoa; Cocoa Products), where the highest number of NPL citations are made to Biological Sciences. Biological sciences, in turn, has a European-to-American ratio below the average, which contributes to the firm receiving a low gap score.

experience). Scientists are assigned a firm-specific weight, as before, to reflect the importance of their scientific field to the firm. We divide European-affiliated scientists by the sum of European-affiliated and non-affiliated scientists to generate an additional AMS-based historical (1921) scientific gap measure.

To calculate the gap measure based on citations to European journals by American journals, we replace $European\ Papers_{field}$ in Equation 1 with the number of citations to European journals made by American journals in that OECD subfield ($American\ Papers_{field}$ is replaced with American citations to American journals). As before, each scientific field is weighted by its relevance to a firm, based on the patent classes the firm patents in and the rate at which patents in the class cite the scientific field. Dividing the European citations by the sum of American and European citations yields the journal citations-based scientific gap measure.

3.4.4 Patents

Our patent data are derived from Google's public patent dataset. There are 637,190 patents granted between 1926 and 1940 by the USPTO. We collect information on the grant date, assignee and inventor names, Cooperative Patent Classification (CPC) codes, as well as prior-art citations made to patents. Based on this, we calculate normalized forward patent citations by dividing the total number of prior-art citations received by a focal patent by the per-patent citations received by all patents granted in the focal patents' issue year.[35] To measure the extent to which a patent "relies" on science, we count citations to scientific publications in Microsoft Academic Graph in the text of the patent, from Marx and Fuegi (2020).[36] We also measure the "novelty" of a patent by counting the number of times the same

[35]This cohort-based normalization is important because a procedural change at the USPTO starting from 1947 substantially increased citations afterwards.

[36]We use in-text citations because NPL citations are available only after 1947. We use references with a confidence score above 8. We find 237 patent citations to science by our sample firms between 1925 and 1940.

CPC combination of a patent has been granted since 1790 (Fleming, 2001).[37] A Combination Familiarity score of zero implies that the technical combination has never appeared before. For instance, Wallace Carother's nylon patent for DuPont (US2130948A) combines eight different CPC subclasses, some for polyamides and others for fibers, which was an unprecedented combination at the time (hence, the resulting familiarity score is zero).[38]

Patent assignees are matched to firms using the same fuzzy string matching algorithm used to match publications.[39] We match 89,328 patents to the 469 firms in our panel between 1926 and 1940. Of these, 234 firms are found in the KKMY sample, 350 are found in CRSP and 115 firms are found in both.

3.4.5 Corporate Ownership, Financial Statements and Industry Concentration

We collect corporate ownership data in order to link firms under common ultimate ownership from KKMY, who collect data on the control of U.S. nonfinancial corporations for the years of 1926, 1929, 1932, 1937, 1940 (and for 1950, after the end of our sample period). Using Moody's Manuals on nonfinancial sectors (Railroads, Public Utilities and Industrials Manuals, which are available from Mergent Online (http://webreports.mergent.com/), the authors start with the largest 200 nonfinancial corporations, ranked by total assets as reported in Berle and Means (1932), and construct ownership trees for these firms, their parents and subsidiaries. The chains of control, from the ultimate owner to all the subsidiaries, are based on Moody's def-

[37] "Combination Familiarity" of patent i, R_i is

$$\sum_k 1\{k \text{ cites identical patent class combination as } i\} \times exp(\frac{\text{publication date of } k \ - \ \text{publication date of } i}{\text{time constant of knowledge loss}})$$

For all patents k granted before i. *time constant of knowledge loss* is set to 5 years.

[38] These are D01F6/60,D01F6/58, D01F6/605, C08G69/26, C08G69/28, D01D5/06, Y10S8/21, Y10T428/2904, Y10T428/2976, and Y10T428/2978.

[39] See appendix B.3.5 for details.

inition of control, which uses both equity links and other considerations. Ultimate owners of control chains (individuals, families or, in some cases, a widely held apex company) are identified using a variety of archival data sources.[40]

Financial Statement Variables — Balance sheet data on earnings and assets are not available before 1950 from conventional sources such as S&P Compustat. Therefore, we build on KKMY, who collect data on firm assets and earnings for the sample firms for the years 1926, 29, 32, 37, 40 (and 50), using Moody's Manuals.[41] We expand this dataset for the intervening years from the same source. To classify the industries in which these firms operate, we use descriptions of firm "occupations" in Moody's Manuals. We then manually connect each industry name to one of the 85 3-digit industry codes in the revised 1947 SIC tables (reported by the BEA in 1958).[42] We augment the dataset by collecting all available end-of-the-year stock market value data for all listed firms using the CRSP Monthly Stock File for North American firms. For listed firms that appear on CRSP but not in the KKMY sample, we obtain data on their financials (but not ownership data) from Graham et al. (2015).[43]

Measures of Industry Concentration — We use Wilcox (1940) to classify 3-digit industries into monopolistic, oligopolistic and competitive ones. Wilcox does not rely on any single test of monopoly or competition to measure the extent of competition in different industries (because of difficulties described in Wilcox, 1940, pp. 1-12; 19-20). Instead, he uses broad criteria, as well as the regional nature of the markets, to separate effectively competitive industries from effectively monopolistic ones as

[40]See Appendix B.3 for details on the construction of ownership chains and the identification of ultimate owners.

[41]Figure B.2 in Appendix B.3.2 reproduces the 1949 entry for the Porto Rico Telephone Company.

[42]Source: http://www.bea.gov/industry/io_histsic.htm.

[43]This dataset is also manually collected from Moody's Industrial Manuals.

of 1934-1939 (the measures do not refer explicitly to a specific year). Monopolistic industries are further divided into cases where supply was predominately accounted for by one or two firms (monopoly and duopoly); by only a few firms (oligopoly); by one or a few dominant firms and many smaller ones (dominant-firm industries) and by several or many firms acting in collusion (cartels and effective trade associations). Industries may be characterized by more than one type of behavior. Industries in which none of these were present were classified as competitive. In our analysis, we focus on the distinction between competitive and non-competitive industries.[44]

3.4.6 Descriptive Statistics

Table 3.2: Summary Statistics of Main Variables

	Observations	Mean	Median	Std. Dev.	Min	Max
Gap in university science (1900-1920)	7035	0.70	0.70	0.03	0.58	0.79
Lab Size	2320	43.49	0.00	214.86	0.00	4669.00
Patents Granted Per Year	7035	13.13	1.00	54.88	0.00	838.00
Publications Authored Per Year	7035	0.46	0.00	3.72	0.00	88.00
Patent Stock	7035	71.18	7.54	316.85	0.00	4441.06
Publication Stock	7035	2.70	0.00	22.16	0.00	440.33
Forward Patent Citations	4035	0.77	0.62	0.80	0.00	18.42
KPSS Patent Value	2629	2.58	1.03	4.37	0.03	56.65
Total Assets ($MM)	4305	1369.71	418.87	3272.52	7.43	60114.66
Gross Income ($MM)	2789	879.27	270.66	1864.59	-1.98	20655.93
Market Capitalization ($MM)	3856	1103.23	248.91	2903.02	0.69	37352.08
Business Group Affiliated = 1	3104	0.41	0.00	0.49	0.00	1.00
Total Assets ($MM) (1926-1930)	1330	1258.17	445.75	2723.14	12.39	36047.36
Gross Income ($MM) (1926-1930)	728	833.95	235.17	1865.36	-1.98	14366.30

Notes: Observations are at the firm-year level, and the sample period is 1926-1940. Forward Patent Citations is first defined at the patent level as the number of forward prior-art citations received normalized by the average number of forward citations for the patent's grant year cohort. This value is averaged at the focal firm-year level to produce the Forward Patent Citations measure in the table. Patent and publication stock are calculated using a perpetual inventory method with a 15% rate of depreciation. KPSS Patent Value is the value of a patent (in million dollars) based on the cumulative abnormal returns in the firm's market value at the issuance event of the patent (Kogan et al., 2017).

Table 3.2 presents descriptive statistics at the firm-year level. The maximal number of observations is 7,035 (469 firms observed for 15 years). "Lab Size" counts the number of lab personnel reported in the IRL directory. There are only around a

[44]Both Stigler (1949) and Nutter and Einhorn (1969) follow, and validate, the Wilcox (1940) classification. Nutter and Einhorn (1969, pp. 94-97) supplement the Wilcox classification using other sources and derive a "concentration ratio" for 1939, albeit only for a subset of sectors.

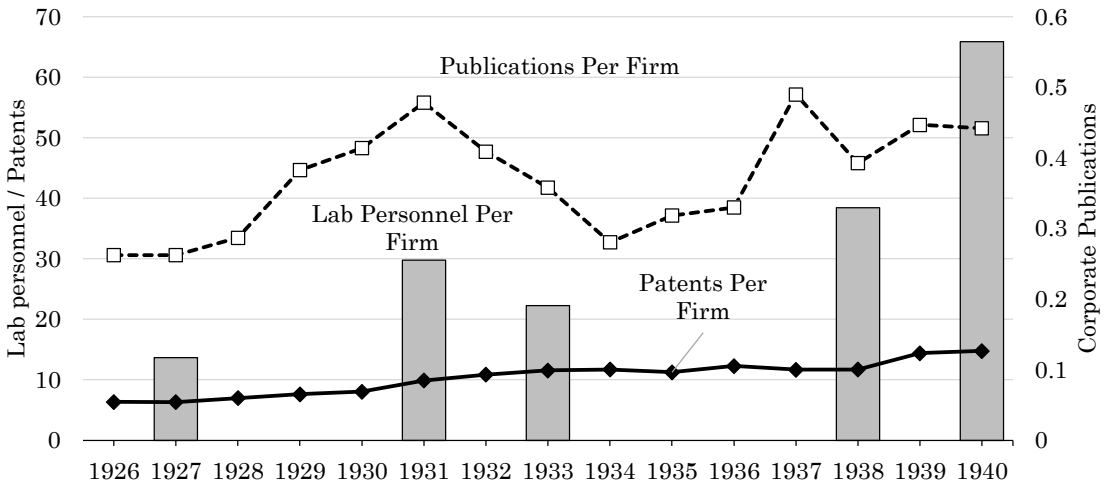


FIGURE 3.3: Emergence of Corporate Science, 1926-1940[45]

third of the total observations here because the IRL was collected for only five years by the NRC (1927, 31, 33, 37, 40). As observed from the difference in median and mean values (see also figure 3.4), scientific publications and lab personnel are skewed to the right. The average gross income and assets are $879 million and $1.4 billion, respectively. These are slightly larger than the values for the "pre-period" between 1926 and 1930, reflecting the growth of the U.S. economy.

Figure 3.3 presents trends in corporate investment in science during the interwar period. Publications per firm exhibit an upward trend, reflecting an aggregate increase from just under 200 corporate-authored papers in 1926 to around 700 papers in 1940. In addition, about 12.5% of all firms published in 1926, increasing to 17.6% of all firms in 1940. The staff employed in corporate laboratories mirrors the same upward trend. These figures reflect an expansion of corporate science both on the extensive margin (measured by total publications) and intensive margin (measured by share of firms that publish). This rise in corporate science, however, is not mono-

[45]The bar graph indicates the number of personnel employed at corporate laboratories per firm from the Industrial Research Laboratories Directory. The broken line indicates the publications per firm in our sample matched to Microsoft Academic Graph. The solid line indicates the number of patents by firms in our sample matched to USPTO utility patents.

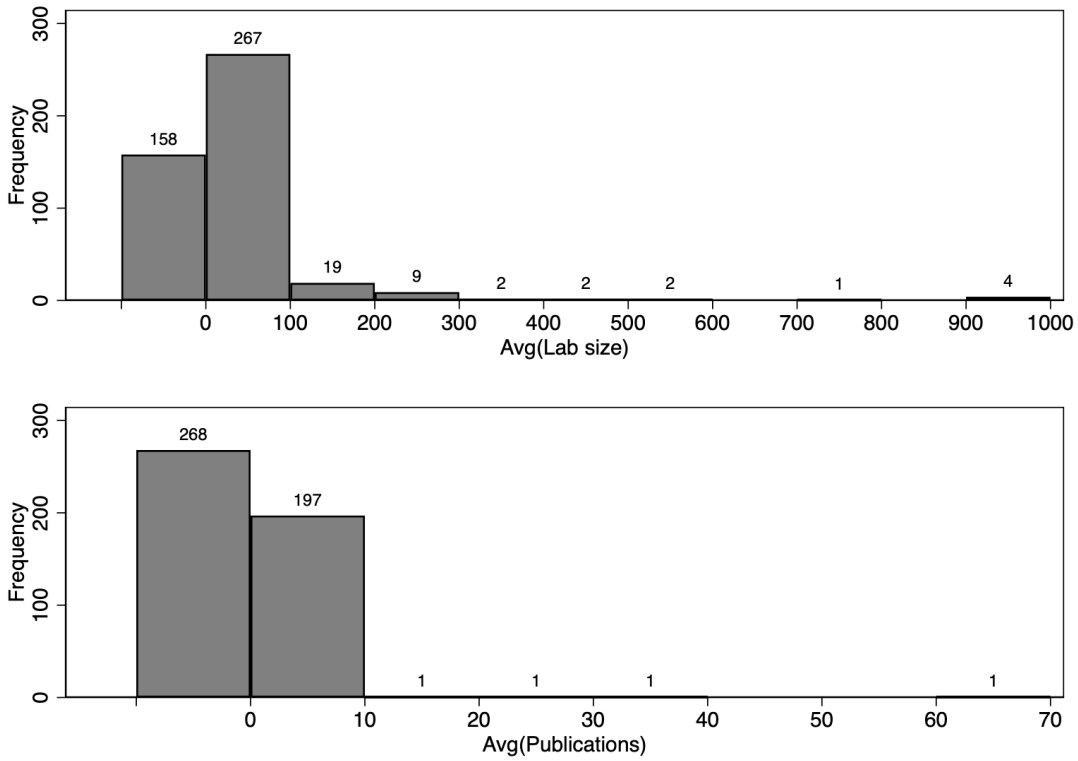


FIGURE 3.4: Heterogeneity of Corporate Science[46]

tonic, peaking in the early 1930s and declining for the next four years, reflecting the cutbacks in research during the Great Depression. The increase in patenting is steadier, if less dramatic, with an aggregate rise from around 3,000 to 6,500 corporate patents per year.

Figure 3.4 shows that these trends mask substantial heterogeneity. For instance, 158 firms out of our sample of 469 firms never operate an R&D lab, while more than half (268) of the firms never publish a scientific article. Perazich and Field (1940) estimate that less than 1% of all firms accounted for a third of all industrial research employment in 1921, 1927 and 1938, respectively. A mere 45 firms in 1938 employed

[46]The upper histogram bins the number of personnel employed at corporate laboratories for firms in our sample. 158 firms (the leftmost bar) report no employed lab personnel in our sample period. The lower histogram bins the number of publications authored by firms in our sample. 268 firms (the leftmost bar) do not author any scientific publications in our sample period.

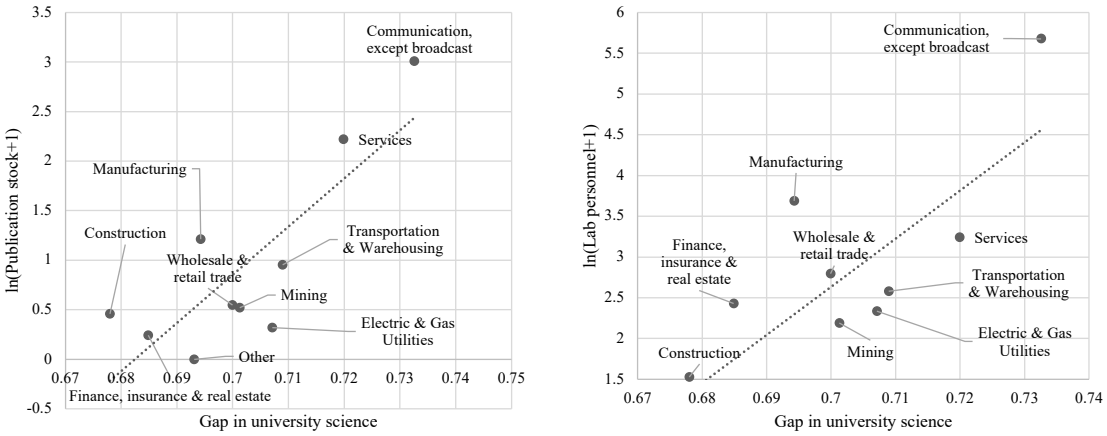


FIGURE 3.5: Corporate Science vs Gaps in University Science, by Industry[47]

half of the total research personnel (Perazich and Field, 1940).

Figure 3.5 presents the correlation between the scientific gap and corporate science across industries.[48] Corporate investments in science seem to occur in industries where the U.S. lags behind European science. For instance, construction, which relies on civil engineering where the gap is small, exhibits less corporate science investment than communications, which relies partly on chemistry, where the gap is large. This pattern is consistent with our conjecture in Section 3.2; it also calls for the use of industry fixed effects throughout the empirical analysis.

3.5 Empirical Results

3.5.1 Who Invests in Science?

We first explore the existence of an extensive margin between "leader" and "follower" firms. If leading firms have higher returns from innovation and there is a fixed cost to corporate research, leaders will invest in research while followers will not. This

[47]Industry-level scatter plots of firm investment in science and the gaps in the relevant academic discipline. The left panel plots the natural log of one plus the average publication stock against the gaps in public science measure. The right panel replaces the publication stock with the number of personnel at R&D labs, from the IRL directory.

[48]We replicate the results with the other two measures in Appendix Figure B.1

accords with the history of corporate science surveyed in Section 3.2: select firms, such as General Electric and DuPont, had reached the technological frontier where the payoffs from fundamental, science-based innovation were higher. Since establishing a research organization to "routinize" such innovation was a costly endeavor, "follower" firms that were not as technologically advanced likely chose not to invest.

We find in the first three rows of Table 3.3 that firms closer to the technological frontier were more likely to publish in scientific journals and operate corporate labs. In particular, firms with patenting intensity greater than the mean tend to have around 22 times greater publication stock and employ around 12 times more lab employees than firms with patent intensity below the mean. Similarly, firms with at least one patent citation to scientific articles during our sample period also tend to have publication stocks that are fiftyfold those of firms whose patents do not cite science, and laboratory personnel fifteenfold larger. Interestingly, these "frontier" firms are in the minority, accounting for barely 6% of the observations.[49]

The other variable heavily associated with investment in research is size. Firms with above average assets publish around 22 times more than those under average, while the difference in lab personnel is around 10 times. Similarly, we find that public firms and firms in the KKMY sample (both correlated with size) tend to invest more heavily in science. Business Group-affiliated firms publish around 16% more publications and employ around 73% more lab personnel.[50] Greater scale may allow firms to reap greater marginal returns from innovation (Cohen and Klepper, 1996). However, it is also possible that firm size and business group affiliation proxies for easier access to internal factor markets that enable conducting basic research (Schumpeter, 1939). Finally, we observe that firms in noncompetitive industries

[49]The total observations for the "Patenting Intensity Above Average?" row (4,305) is smaller than the "Patents Cite Science" rows (7,035) due to missing sales data.

[50]However, unlike the previous results, the difference between business group affiliates and non-affiliates is not statistically significant at the 5% level.

tend to engage more in corporate science (more publications and larger corporate laboratories). This is consistent with the conjecture that firms in concentrated or non-competitive industries, may be able to capture the rents from the provision of scientific knowledge, a quasi-public good.

Table 3.3: Investment in Corporate Science, by Firm Characteristics

		Avg(Publication Stock)		Avg(Lab Personnel)		Obs
Patenting Intensity Above Average?	Yes	16.89	(2.22)	224.64	(31.78)	768
	No	0.75	(0.04)	19.41	(1.55)	3537
Patent Forward Cites Above Average?	Yes	5.82	(0.66)	89.70	(10.55)	2788
	No	0.65	(0.04)	11.28	(1.17)	4247
Patents Cite Science?	Yes	32.20	(3.86)	343.50	(57.90)	450
	No	0.68	(0.03)	22.75	(1.94)	6585
Public?	Yes	3.37	(0.35)	47.19	(5.29)	5250
	No	0.73	(0.05)	32.64	(8.18)	1785
KKMY?	Yes	4.90	(0.53)	73.30	(8.87)	3510
	No	0.51	(0.03)	14.45	(1.22)	3525
Assets Above Average?	Yes	7.38	(0.87)	109.46	(13.20)	2014
	No	0.34	(0.03)	11.28	(1.09)	2291
Business Group?	Yes	5.35	(1.41)	175.29	(67.69)	477
	No	4.63	(1.20)	101.07	(23.15)	693
Competitive Industry?	Yes	1.12	(0.10)	28.68	(3.07)	2865
	No	3.89	(0.45)	54.85	(7.40)	4050

Notes: Unit of analysis is the firm-year. Standard errors are indicated in parentheses. "Patenting Intensity" is defined as number of patents granted divided by log of assets. "Patents Cites Science" is equal to "Yes" if the firm's patents cite at least one scientific article and "No" otherwise (data from Marx and Fuegi (2020)). Observations with zero patents also enter the "No" category. "Public?" and "KKMY?" respectively ask whether the firm is a listed firm found in the CRSP dataset and the (Kandel et al., 2019) dataset from Section 3.4. Observations enter the "Yes" row for "Assets Above Average?" if the total assets of the observations are above average. "Business Group?" asks whether firms are part of ownership chains with more than three firms per the definition in Kandel et al. (2019). "Competitive Industry?" rows classify firms based on their industry's competition classifications in Wilcox (1940); firms in industries classified as oligopolies and monopolies are classified in the "No" group.

3.5.2 Interaction with Gaps in University Science (OLS)

We next explore whether gaps in public science accentuate the incentives of those leading firms to invest further in research. A concern is that our findings are driven by unobserved heterogeneity in firm quality. Unobserved firm quality should be pos-

itively correlated with technological leadership and internal science, leading to an upward bias in their estimated relationship. To alleviate this concern, we exploit variation across technology fields differing in the availability of public science. Our theory predicts that technology leaders should find it attractive to invest in internal science, especially when relevant public science is relatively scarce. When public science is scarce, leading firms, which benefit more from science, have a strong incentive to invest. Yet, as public science becomes abundant, science is no longer a source of private value, and investment in internal science becomes less attractive. Therefore, using public science as a moderator of the relationship between technology leadership and internal science, and showing that our key results hold when public science is scarce, should alleviate, at least partly, the unobserved firm quality concern.

We estimate the following specification via OLS, where we measure the availability of public science using our scientific gap measure(s):

$$Corpsci_{it} = \beta_0 + \beta_1 MR_Inno_i + \beta_2 Gap_i \times MR_Inno_i + \beta_3 Gap_i + \boldsymbol{Z_i'\gamma} + \boldsymbol{\tau_t} + \boldsymbol{\phi_c} + \epsilon_{it} \tag{3.3}$$

where $Corpsci_{it}$ is defined as the investment in corporate science by our sample firms, measured by the scientific publication stock;[51] MR_Inno_i refers to the marginal returns to innovation for firms. One proxy for these returns is proximity to the technological frontier (need for science): i) patent intensity (patents granted divided by log of total assets), ii) average forward citations the firm's patents receive and iii) the number of citations to scientific publications the firm's patents make. We also replace MR_Inno with firm characteristics indirectly related to returns to innovation: i) the natural logarithm of the firm's total assets; ii) a dummy variable denoting whether the firm is part of a multi-firm business group and iii) a dummy variable

[51] Using number of lab employees yields similar results. However, with fewer than half the observations, the estimates are far less precise.

denoting whether the firm's industry is classified as "competitive" by Wilcox (1940). Gap_i is measured as the ratio of European publications to American publications relevant to each firm[52]; $\boldsymbol{Z_i'}$ is a vector of controls. We include level values for each science measure since our gap measure is a ratio. We also control for the size of the firm (assets) and include year (τ_t) and 2-digit industry (ϕ_c) fixed effects. We expect $\hat{\beta}_2 > 0$ if the gap in public science accentuates "leader" firm investment in science.

Table 3.4: Corporate Publications and Gaps in University Science, by Proximity to Technological Frontier

	DV: Publication stock		
	(1)	(2)	(3)
Gap in university science, 1900-20 × Patenting Intensity	99.098		
	(27.004)		
Gap in university science, 1900-20 × Forward Patent Citations		45.762	
		(11.949)	
Gap in university science, 1900-20 × ln(Patent Cites to Science)			1466.288
			(220.445)
Gap in university science, 1900-20	-7.364	64.508	66.172
	(9.829)	(16.099)	(14.400)
Patenting Intensity	-64.052		
	(19.291)		
Forward Patent Citations		-29.700	
		(8.143)	
ln(Patent Cites to Science)			-994.021
			(156.863)
ln(Assets)	-0.352	4.168	2.842
	(0.248)	(0.584)	(0.429)
ln(American pubs), 1900-20	-0.348	2.615	2.239
	(0.580)	(1.016)	(0.796)
ln(European pubs), 1900-20	1.160	0.214	0.238
	(0.556)	(0.880)	(0.645)
Average of Dependent Variable	3.642	3.642	3.642
Year Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
R^2	0.683	0.151	0.291
Number of Firms	422	422	422
Number of Observations	4,293	4,293	4,293

Notes: Unit of analysis is at the firm-year level. "Gap in university science, 1900-20" is defined at the firm level as the number of European scientific publications relevant to a firm divided by all (European & American) publications relevant to a firm (winsorized at the 1% and 99% level). "Patenting Intensity" is defined as number of patents granted divided by log of assets. "Patent Cites to Science" counts the number of in-text citations made by the focal firm's patents to scientific publications in MAG. "Business Group Dummy (Imputed)" refers to firms that are Business Group affiliates (defined as ownership chains with at least three firms). "ln(Assets) (1926-1930)" takes the natural log of average assets between 1926-1930. "Competitive market" is a dummy equal to one if the firm's industry is classified as "competitive" by Wilcox (1940). The rest of the variable definitions are identical to those in Table 3.2. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

[52]We present results using the two alternative measures in Appendix B.2

Table 3.4, Columns 1-3 show that firms with higher patenting intensity, whose patents receive more forward citations and whose patents cite the scientific literature were more likely to respond to a gap in university science. Specifically, firms with patents in the 75th percentile of forward patent citations respond around 11% more in terms of new scientific publications compared to firms whose patent citations are in the 25th percentile. We also find similar results when normalizing firm scale (e.g., dummy for whether the firm patents cite science, number of patent citations to science normalized by number of firm patents), but we prefer to independently control for size using assets to also understand the baseline relationship between investment in science and size. These results are consistent with our conjecture in Section 3.2 that, on average, more technologically advanced firms experienced a more acute need for internal research when faced with weaker domestic universities. In terms of the theoretical framework, these findings suggest that university science was a substitute for internal research or that there are strong strategic complementarities in innovation, or both.

Table 3.5: Corporate Publications and Gaps in University Science, by Size and Corporate Structure

	DV: Publication stock		
	(4)	(5)	(6)
Gap in university science, 1900-20 × ln(Assets) (1926-1930)	140.875		
	(23.226)		
Gap in university science, 1900-20 × Business Group Dummy (Imputed)		282.032	
		(81.008)	
Gap in university science, 1900-20 × Competitive market			-139.601
			(22.996)
Gap in university science, 1900-20	-3614.816	31.827	132.589
	(601.497)	(47.901)	(24.099)
ln(Assets) (1926-1930)	-94.513		
	(15.771)		
Business Group Dummy (Imputed)		-198.801	
		(56.328)	
ln(Assets)	0.392	6.599	4.326
	(0.643)	(1.081)	(0.601)
ln(American pubs), 1900-20	2.354	6.334	3.278
	(1.303)	(3.197)	(1.050)
ln(European pubs), 1900-20	0.513	0.160	-0.362
	(1.193)	(2.779)	(0.888)
Average of Dependent Variable	3.843	6.625	3.657
Year Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
R^2	0.183	0.180	0.152
Number of Firms	358	199	420
Number of Observations	4,028	1,702	4,275

Notes: Unit of analysis is at the firm-year level. "Gap in university science, 1900-20" is defined at the firm level as the number of European scientific publications relevant to a firm divided by all (European & American) publications relevant to a firm (winsorized at the 1% and 99% level). "Patenting Intensity" is defined as number of patents granted divided by log of assets. "Patent Cites to Science" counts the number of in-text citations made by the focal firm's patents to scientific publications in MAG. "Business Group Dummy (Imputed)" refers to firms that are Business Group affiliates (defined as ownership chains with at least three firms). "ln(Assets) (1926-1930)" takes the natural log of average assets between 1926-1930. "Competitive market" is a dummy equal to one if the firm's industry is classified as "competitive" by Wilcox (1940). The rest of the variable definitions are identical to those in Table 3.2. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

We also examine interactions between size and scientific gaps in table 3.5. Column 1 shows that the publication sensitivity to the gap in science of firms with a one standard deviation larger-than-the mean assets is around 1.2 times the estimated sensitivity of a firm whose assets equal the sample mean. This suggests that larger firms were more likely to internalize the effects of corporate research.[53] Similarly, firms

[53]In unreported robustness checks, we replicate this result by replacing the continuous measure of size with a dummy equal to one if a firm has assets above the sample average, finding similar results

affiliated with business groups published about 9 times more in response to gaps in university science (Column 5). We test whether firms operating in more concentrated (less competitive) industries are willing to invest more in science. Other things equal, firms in industries with higher concentration may be less concerned about spillovers, since they can more easily appropriate the returns to their research. Consistent with this, Column 6 shows that being in a competitive industry is negatively correlated with engaging in corporate research in response to scientific gaps.

We find that the non-interacted, standalone coefficients measuring technological frontier and firm size are negative (for instance, -64.052 for patenting intensity and -94.513 for assets). However, the inflection points for the marginal "effects" of these variables occur at fairly low values of the scientific gap measures. For instance, the marginal difference in publication stock related to patenting intensity becomes positive starting from a scientific gap of 0.65, which is in the lower fifth percentile of this measure.[54]

We conclude that firms most likely to respond to gaps in science are the technologically advanced firms. Firm size and organization and, in particular, affiliation with business groups and other multi-firm entities, are related to the likelihood of engaging in research, presumably because of the ability of such corporate structures to internalize the benefits generated by the creation of basic knowledge.

3.5.3 Performance Consequences of Corporate Science

Nelson (1959a, p.119) noted that "Research laboratories may be created and maintained by firms for many purposes, including (...) quality control, (...) improvement of manufacturing methods, improvement of existing products and development of new uses for them, development of new products and processes, and *scientific re-*

[54]In unreported robustness checks, we replicate the results with only level coefficients (excluding interaction terms, but including identical controls and fixed effects) to find that the average marginal effects on publication stock of these variables are positive, consistent with Table 3.3.

search to acquire knowledge enabling more effective work to be done to achieve the above purposes" (emphasis added). If so, firms investing in scientific research ought also to have better and more valuable new inventions. We use two measures: patents deemed valuable by investors (i.e., whose issuance is associated with increases in the firm's stock price), and novel patents, which combine patent classes which have rarely been combined before. As a benchmark, we first estimate the following OLS specification, which includes firm publications as a proxy for corporate science and controls for the current (contemporaneous) WoS gap measure:

$$HomeRun_{it} = \beta_1 Corpsci_{it} + \boldsymbol{Z_i'\gamma} + \boldsymbol{\tau_t} + \boldsymbol{\phi_c} + \epsilon_{it} \tag{3.4}$$

$HomeRun_{it}$ is measured using the number of patents that are in the top 5% of stock market value (Kogan et al., 2017) and novelty scores (Fleming, 2001).[55] One possible concern is that the OLS estimate of β_1 may be upward biased if unmeasured technical opportunity drive both corporate publication activity and the value (or novelty) of patents. On the other hand, it is possible that other firm characteristics, such as the quality of the firm, may be negatively correlated with the current technical opportunities, leading to a downward bias. That is, it is possible that technical opportunity is reflected in greater public, not private, science. We instrument for corporate publications using historical public science gap, which can also purge the measurement error in using publication stock as a measure of investment in corporate research.[56]

Our instrument is calculated using American and European scientific publications authored before our sample period.[57] Historical gaps that predate the sample period

[55]Results are not sensitive to the use of alternative thresholds, such as top 1% or 10%.

[56]Highly novel patents are correlated with patent quality (patent intensity, forward citations, citations to science used in Table 3.4, Columns 1-3), since both are patent-based measures. Because higher patent quality is also correlated with higher publication stock (per first three rows of Table 3.3), this may lead to a spurious relationship between publication stock and patent novelty.

[57]Our sample period is between 1926 and 1940, and the publications for the gaps collected from

affect investment in science but are unlikely to affect concurrent firm inventions and value. In addition, patent value and inventive activity might be affected by the concurrent availability of public science. Therefore, we include concurrent gap measures as controls. We use two-stage least squares, where in the first stage, we regress $Corpsci_{it}$ on the pre-sample period public science gap measure (from Equation 3.3) and other controls, and in the second stage we regress $HomeRun_{it}$ on the fitted values of investment in corporate science ($\widehat{Corpsci}_{it}$) obtained in the first stage.

Market value data for patents are only available for public firms; nevertheless, as shown in Table 3, the bulk of corporate research is carried out by publicly traded firms. Indeed, public firms account for the majority of firms in our sample and constitute the vast majority (around 80%) of firms that publish or operate labs.

1900 to 1920. Similarly, the AMS directory data for the gap measure calculation is from the 1921 edition, while the journals used for the citation share gap measure are those published between 1900 and 1920.

Table 3.6: Corporate Science and "Home-Run" Patents

	DV: Top 5% Xi			DV: Top 5% Novelty		
	(1) OLS	(2) 1st Stage IV	(3) 2nd Stage IV	(4) OLS	(5) 1st Stage IV	(6) 2nd Stage IV
Publication Stock (100s)	0.040		0.137	0.273		0.484
	(0.012)		(0.027)	(0.020)		(0.051)
Gap in university science, 1900-20		95.653			74.670	
		(15.705)			(13.514)	
Gap in university science, current	-2.728	-5.580	-1.832	0.974	-2.219	1.926
	(1.103)	(5.115)	(0.984)	(1.627)	(4.079)	(1.585)
ln(Assets)	1.419	2.529	1.188	0.761	2.408	0.276
	(0.218)	(0.471)	(0.202)	(0.160)	(0.400)	(0.153)
ln(Patent stock)	0.589	5.222	0.094	1.866	4.204	0.989
	(0.081)	(0.563)	(0.091)	(0.152)	(0.477)	(0.156)
Average of Dependent Variable	0.864		0.861	2.903		2.895
Kleibergen-Paap rk Wald F statistic		37.097			30.530	
Year Fixed Effects	Yes		Yes	Yes		Yes
Industry Fixed Effects	Yes		Yes	Yes		Yes
R^2	0.248		0.023	0.635		0.378
Number of Firms	327		327	425		425
Number of Observations	3,569	3,569	3,569	4,293	4,293	4,293

Notes: Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan et al., 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5 and 6 is the share of European papers ("Gap in university science, 1900-20") for papers published between 1900 and 1920. "Gap in university science, current" calculates the share of European publications in the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

OLS estimates from Table 3.6, Columns 1 and 4, show that firm publication stock is positively correlated with the number of highly valuable (within top 5% of stock market value per Kogan et al. (2017)) and highly novel (within 5% of novelty scores measured by the count of subclass combinations per Fleming (2001)) patents. In Columns 2 and 5 we instrument for firm publication stock using the historical (pre-sample period) gaps in university science. The first stage regressions of publication stock against gaps in university science are significant, with an F-stat of 37. Because American universities were catching up to European standards during the interwar period, we also control for the "current" gaps calculated for each firm and year and find it to be negatively correlated with the corporate publication stock.[58] In the second stage regressions (Columns 3 and 6), we find that an increase in the (instrumented) publication stock increases the number of valuable patents. A one standard deviation larger publication stock (due to historical gaps in university science) leads to around three more patents in the top 5% of stock market value (or around 3.5 times the sample mean). The estimate in Column 3 is about 3.4 times larger than the OLS estimate in Column 1 (similarly, the estimate in Column 6 is 1.8 times larger than OLS estimate in Column 4). One interpretation is that the corporate investments in research undertaken when public science lags are particularly potent sources of competitive advantage, as reflected in more valuable and distinct inventions. Put differently, publication stocks might have heterogeneous effects on invention outcomes. Publication stocks that reflect investments in response to gaps appear to have a larger impact on invention outcomes. We find similar results in Appendix Tables B.3 and B.4 using the two alternative measures of public science gaps between America and Europe. In sum, investments in internal research do result in

[58]The $European\ Papers_{subfield}$ in the context of Section 3.4.3 are now calculated for each year, instead of summed up for papers published between 1900 and 1920. The mapping from scientific field to firm are identical, and the gap measure is calculated at the firm-year level, not the firm level.

more valuable inventions.

As an alternative to estimating "technological" (patent-based) returns, we also estimate the contribution of corporate research to firm value of public firms:

$$ln(Q)_{it} = \beta_1 ln(Pubstock_{it-1}) + \boldsymbol{Z_i'\gamma} + \boldsymbol{\tau_t} + \boldsymbol{\phi_c} + \epsilon_{it} \tag{3.5}$$

where Q is the market-to-book ratio (or Tobin's Q) and $ln(Pubstock_{it-1})$ refers to the natural log of one plus lagged publication stock.

Table 3.7: Corporate Science and Market-to-Book Ratios

	Baseline (OLS)	Gap Split (OLS)		IVE	
	(1)	(2) Small	(3) Large	(4) 1st Stage	(5) 2nd Stage
ln(Pubstock$_{t-1}$)	0.032	-0.001	0.045		-0.040
	(0.008)	(0.019)	(0.009)		(0.081)
Gap in university science, 1900-20				2.867	
				(0.492)	
ln(Patstock$_{t-1}$)	0.023	0.042	0.010	0.222	0.038
	(0.004)	(0.007)	(0.006)	(0.013)	(0.018)
Gap in university science, current	-0.177	0.084	-0.204	-0.021	-0.187
	(0.083)	(0.153)	(0.103)	(0.186)	(0.082)
Average of Dependent Variable	0.591	0.607	0.575		0.591
Kleibergen-Paap rk Wald F statistic				28.916	
Year Fixed Effects	Yes	Yes	Yes		Yes
Industry Fixed Effects	Yes	Yes	Yes		Yes
R^2	0.346	0.400	0.396		0.003
Number of Firms	325	170	155		325
Number of Observations	3,399	1,740	1,659	3,399	3,399

Notes: Unit of analysis is at the firm-year level. Columns 1-3 present results from estimating the Tobin's Q equation against lagged publication stock. Columns 2 and 3 are split by mean values of the "Gap in university science, 1900-20" measure (2 being below average and 3 above). Column 4 presents the first stage IV estimates, where lagged publication stock is predicted by share of European publications published between 1900 and 1920 ("Gap in university science, 1900-20"). Column 5 regresses Tobin's Q against the predicted lagged publication stock from column 4. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Column 1 of Table 3.7 shows that publication stock is positively related to the market-to-book ratio of firms. A one standard deviation larger publication stock is associated with an increase of Tobin's Q by around 0.01 (2% of the sample mean). We also split the sample by gap measures to probe whether "responding firms" benefited more from science. Comparing Columns 2 and 3, we find that, for firms whose gaps in university science (based on author affiliations) are smaller than the sample mean,

there is a statistically insignificant correlation between their publications and their market-to-book ratios. In contrast, for firms with gaps above the sample mean there are positive and significant effects of the publication stock on Q. These results suggest that investments in science are positively related to market value, and that this relationship is driven by firms with large scientific gaps.

We also instrument publication stock by the gap in university science. As expected, we find in Column 4 that the first stage regression coefficient of publication stock against gaps in university science is positive and statistically significant. However, we do not find significant results for the second stage. This may be due to the fact that the market-to-book ratio is affected by many firm-level characteristics and is imprecisely measured in the historical data, making it difficult to estimate the effect of corporate research on it. When we look at a more narrow definition of financial returns, we find that the stock market value of patents is positively related in the second stage to publication stock predicted by our scientific gap measures (Column 3 of Table 3.6).

3.6 Discussion and Conclusion

We argue that the rise of corporate research in America in the interwar period is related to the weakness of American academia in certain scientific fields. For some firms - large, group affiliates and close to the technological frontier - investment in internal research was the way to overcome this institutional weakness, or void, as well as a source of competitive advantage: competitors would not be able to readily acquire the needed scientific knowledge from universities.

This historical evidence on research carried out by private corporations may be of relevance to the present-day debate in the U.S. and other advanced economies about the costs and benefits of large, technologically advanced firms. Our results suggest that such firms can play an important role in advancing knowledge, but that this

knowledge is likely to grant them considerable advantage over their competitors. In emerging economies, our historical analysis sheds light on a relatively little-explored mechanism by which large corporate entities attempt to make up for institutional voids, in this context, voids related to the accessibility of science and the quality of domestic academic research. Our historical evidence suggests that, in some circumstances, private corporations can and do substitute for institutional weaknesses in science, as some corporate giants in contemporary emerging markets (e.g., India, Turkey and Korea in its early stages of development) have done. Naturally, such historical parallels should be used with caution.

In the decades since the end of our sample period, American universities have increased the quality and quantity of their scientific output, yet the implications of this change for corporate research remain poorly understood. As noted at the outset, universities produce both new knowledge as well as human capital, which affect both the costs and benefits of private investment in research and innovation. Following World War-II, the growth of university research was paralleled by growing investments in corporate research. However, by the 1980s, the two trends have diverged, leading to a growing division of labor between academia and universities (Arora et al., 2018b).[59] How the private value of an input changes as the supply of the input expands is an important but understudied topic. When the input in question is knowledge, whose use by one firm does not preclude its use by another but may affect the private value that accrues to the firms, the issue becomes even more complex.

In addition to providing new evidence on corporate research in America in the interwar period, we assemble the most extensive historical sample of American firms

[59]Besides the growing scientific might of American universities, it is possible that corporate research in the 1930s was easier protect from rivals (few could effectively use it), whereas by the 1980s knowledge spillovers may have become increasingly costly. This conjecture is consistent with Arora et al. (2021b), who show that companies cut back on research when spillovers to rivals increased relative to the value from internal use.

that were involved in innovation during that period, including information on the scientific output of these firms and on the relative gap between American and European universities. We hope that these newly developed data will contribute to future research on the open questions we raised.

4

Science, Startups and Novelty

4.1 Introduction

This paper asks why startups are more likely to introduce novel innovations utilizing new science than incumbents. I put forward a theory that explains startups' higher propensity to engage in novel innovations by distinguishing between pre-entry invention and post-entry innovation. The former corresponds to conceiving new ideas or processes; the latter corresponds to commercializing these in the market. In this conceptualization, startups have an absolute commercialization disadvantage relative to incumbents who benefit from prior investments in manufacturing, sales, and distribution. However, this disadvantage incentivizes startups to experiment with novel innovations that may be riskier but offer higher payoffs upon success. This implies that startups should enter the product market less often conditional upon patenting, and commercialize more novel products conditional upon entering. I test this theory in the context of pre-entry patenting and post-entry product introductions that occurred in the American laser industry due to access to Soviet quantum electronics knowledge made abruptly available after the Cold War. This setting creates an even

playing field for startups and incumbents in the pre-entry period of invention and helps address concerns that the observed differences in post-entry product novelty are confounded by some firms being inherently better during the pre-entry phase.

Novel laser product introductions linked to areas of Soviet scientific excellence increase two-fold after the end of the Cold War. Consistent with the prior literature, I find that startups are around 1.5 times more likely than incumbents to sell novel products using Soviet science after the Cold War. This difference is greater for lasers with new applications (and muted for those without). Startups using Soviet science are also less likely to exit from the product market than incumbents. Diverging from the literature on capabilities, I find that the startup "advantage" post-entry is not mirrored in the pre-entry stage during invention: startups are no more likely than incumbents to cite Soviet journals or generate novel patents using Soviet science. Also, startups are only around half as likely to enter the product market compared to incumbent firms conditional upon patenting. These patterns strongly suggest the role of selection into product market entry: startups do not invent more novel lasers pre-entry, but those that enter do so with more novel products that can offer higher returns. This is consistent with the startup-incumbent parity I find during pre-entry patenting, lower entry rates of startups upon patenting, higher share of novel products for startups upon entry, and lower exit afterwards. To address endogeneity concerns that the reallocation of Cold War era defense spending on laser R&D in America may bias my results, I replicate the results after excluding lasers that were related to the Strategic Defense Initiative (SDI), which began during the Reagan administration.

I contribute to the literature that has examined what kinds of firms successfully respond to "technological discontinuities" (Tushman and Anderson, 1986). Earlier works such as Chatterjee and Wernerfelt (1991) and Montgomery and Hariharan (1991) theorized that prior experience provides incumbents with better "upstream"

capabilities that enable the firm to enter into related markets.[1] The subsequent literature has provided empirical validations, showing that prior experiences in the television (Klepper and Simons, 2000),[2] fiber optics (Cattani, 2005), hard disk drive (King and Tucci, 2002; Franco et al., 2009), and flat panel display (Eggers, 2014) industries allow incumbents to rapidly adapt to new technological developments. But counterexamples also persist, most notably in the photolithography (Henderson and Clark, 1990) and hard disk drive industries (Christensen, 1993), where incumbents often fail to adapt to new technology. A review of the literature concludes that the evidence on startup vs incumbent advantages in new technology adaptation is mixed (Ganco and Agarwal, 2009) and later work has qualified that prior experiences can sometimes create cognitive biases (Benner and Tushman, 2002; Gilbert, 2005) and function as "prisms" that distort responses to new technologies as much as "pipes" supplying them (Wu et al., 2014).

Still, even if one assumes firms *can* enter because of superior upstream capabilities, it does not follow that they *will* do so, as innovation payoffs also depend on downstream market conditions. Indeed, technologically advanced firms in the pre-entry "incubation" phase of a new technology often sell or license their inventions to downstream commercializers instead of entering the market themselves (Moeen and Agarwal, 2017). The technology commercialization (Teece, 1986; Gans et al., 2002) and division of innovative labor literature (Arora et al., 2004; Gans and Stern, 2010) also theorize that downstream appropriability conditions can deter startup entry. Specialized complementary assets in the medical diagnostics (Mitchell, 1989), PC, steel, pharmaceuticals, and telecommunications (Rothaermel and Hill, 2005), and

[1] "Upstream" capabilities refer to the ability to initiate or absorb new technological developments (the "invention" phase), while "downstream" capabilities refer to economic value capture (the "innovation" phase).

[2] Klepper and Simons (2000) find that radio producers' prior experience allowed them to survive longer than non-radio producers as the television industry was emerging.

typesetter (Tripsas, 1997) industries were found to increase entry for incumbents in new technologies, even if they were "competence-destroying".

Overall, the prior literature affirms the intuition that entry costs affect whether and when a firm enters, but does not ask how the novelty of innovations varies by the type of entrant, which is the gap I fill with my study. Rothaermel and Hill (2005) show returns on earnings and on assets were higher for incumbents in the pharma and telecomms industries following advances in recombinant DNA technology and wireless telephony, but not for the computer and steel industries after the advent of the PC and electric arc furnace. However, since data is limited to public firms, innovativeness of *de novo* entrants that were likely smaller (and thus unlisted) is only described anecdotally. Mitchell (1989) compares the probability of entry among incumbents into five new generations of medical imaging technology to find that incumbents with complementary assets enter more and earlier, but does not examine why startups (*de novo* entrants) account for 81% of all entrants into these new subfields.[3] Tripsas (1997) shows that the market share of *de alio* entrants with specialized downstream assets is higher even though technical performance (measured by newspaper lines printed per minute) is higher for *de novo* entrants across all generations of typesetters (analog, digital CRT, and laser typesetters). While this fact is used to underline the importance of downstream assets in "buffering" incumbents against technological change, why novel startup typesetters were superior in the first place is left unexplained.

Attention to the entry threshold mechanism offers a fresh perspective on several patterns in the literature where startups seem to out-innovate incumbents (Acs and Audretsch, 1987; Griliches, 1990). For instance, Kapoor and Furr (2015) show that

[3]From the descriptives in Table 1 of the paper, incumbents account for only 58 of the 320 total entrants into the five new subfields. The incumbent share also varies widely, from 7% (12 out of 165) in ultrasound to 62% (18 out of 29) in digital radiography. It is unclear whether the entry rates are high or low for the startup firms because the paper's data does not identify potential startup entrants (via the list of patentees, for instance).

startup entrants in the photovoltaic (PV) industry offset their weaker downstream capabilities with superior technology. The authors attribute this to differences in the pre-entry downstream capabilities between startups and incumbents: incumbents specialize in PV types where there are more complementary assets while startups that lack these assets specialize in increasing technological performance. However, the reasoning is still largely based on the assumption that pre-existing assets offer higher returns on further investment. My theory instead explains this specialization through startups' needs to overcome higher entry thresholds imposed by the incumbents. Neither the argument that selection contributes to the types of entrants (Thompson, 2005; Chang and Wu, 2014) nor the argument that downstream market conditions affect the types of innovations (Schmookler, 1962) are new. However, testing my theory has been empirically difficult because startups and incumbents differ in their pre-entry "upstream" endowments, as the literature above shows.

Therefore, I find a setting where the "upstream" science is new to both startup and incumbent firms: the American laser industry after the end of the Cold War, which unexpectedly increased access to Soviet laser science. I then link the full "supply chain" of innovation from science (publications) to invention (patents) and innovation (products) through the chemical compounds of the optical gain media used in lasers.[4] Linkages between technical attributes and products have been studied in flat panel displays (Eggers, 2014), photolithography (Henderson, 1992), and photovoltaic cells (Kapoor and Furr, 2015), but these do not stretch back to the underlying science. Moreover, the precise "risk set" of inventors pre-entry and the innovators post-entry have not been distinguished in these studies.[5] This has been pointed out as a limitation in the empirical literature comparing entrant and startup

[4]Please see Section 4.4 for details on the linkage.

[5]Mitchell (1989) and Tripsas (1997) both explicitly consider a "risk set" of potential entrants, but this set is limited to firms that field at least one product, which omits inventors that never enter.

survival advantages (Bayus and Agarwal, 2007; Lourdes Sosa, 2013), since there are often more firms exploiting a new technological development than those entering the downstream market (Moeen and Agarwal, 2017). By linking laser types across scientific papers, USPTO patents, and products sold in industry buyers' guides, this paper fills this empirical gap in the literature. On the other hand, studies exploiting geopolitical shocks to the supply of scientific knowledge (Ganguli, 2015; Borjas and Doran, 2012; Moser et al., 2014; Iaria et al., 2018; Teodoridis et al., 2019; Ferrucci, 2020) have primarily focused on the impact on scientific knowledge itself, and therefore have been silent on the knock-on effects on innovation. Moser et al. (2014) is an exception which studies follow-on effects on chemical patenting due to the immigration of German scientists fleeing the Third Reich. However, this study does not examine how patenting was related to entry into chemical product markets. Ferrucci (2020) studies differences in German patenting patterns with the fall of the Berlin Wall, but does not link these to product market outcomes. The linkage between Soviet science and select laser inventions and products (and firm characteristics) hence constitutes an empirical advance on prior works.

The rest of the paper is organized as follows: Section 2 sets the theoretical framework of the paper; Section 3 explains the empirical context of the laser industry and Soviet laser science; Section 4 explains the sourcing of the data and construction of key variables; Section 5 includes the results; Section 6 concludes.

4.2 Theoretical Framework: Science and Novelty of Innovation

Since the work of Evenson and Kislev (1976), the trade-off between novelty and success in technological search has been well established: searching novel products has higher value but lower probability of success; pursuing incremental improvements on the other hand has lower value but higher chances of success. Search without scientific knowledge amounts to "local search", relying on trial and error. But science

provides a "map" that focuses this search toward more efficient avenues (Rosenberg, 1974; Nelson, 1982; Fleming and Sorenson, 2004). The search for vaccines against SARS-CoV-2, for instance, has been hastened by the rapid sequencing of its genome and prior research on protein spikes.[6] Indeed, recent research using corporate drug trials shows that publicly funded cancer gene mapping data also increased the success rate of private research (Kao, 2019).

A relatively understudied question, however, is how startups and incumbents respond to the lower costs of technological search. Indeed, according to Cohen (2010, p.137), "Whether new ventures and entrants...are chiefly responsible for "radical" innovation — though often talked about — suffers from a dearth of rigorous empirical study." Faced with a scientific advance, firms can choose between improving approaches to existing problems (incremental) or solving new problems altogether (novel). The former may be safer but less profitable (lower mean and variance in payoffs) compared to the latter, which is riskier but more profitable (Ahuja and Morris Lampert, 2001; Gatignon et al., 2002; Fleming, 2001). However, there is surprisingly sparse empirical research within industries that distinguishes the types of products (novel vs incremental) introduced when comparing startups and incumbents. Prusa and Schmitz Jr (1991)'s study of innovations across 17 software categories is an exception that shows startups to create new software categories, while incumbents develop existing ones. However, this study neither examines firm invention distributions pre-entry nor offers a theoretical explanation for this regularity.

Anecdotal evidence from the laser industry suggests that startups were more willing to experiment with commercially less certain but potentially more profitable opportunities. For instance, research on optical fibers led to the discovery that fibers can be doped with rare earth elements to be used as gain media for lasers. Incum-

[6]https://www.nytimes.com/interactive/2021/03/25/magazine/genome-sequencing-covid-variants.html

bents in this area such as AT&T extended this technology to address existing needs in fiber networks: their largest application of this technology was in Erbium Doped Fiber Amplifiers (EDFAs) that amplify optical signals over extended distances for fiber networks. Startups such as IPG, on the other hand, invented a completely different product (a high-powered fiber laser) that could be used in industrial marking operations, where high power was valuable, unlike in telecommunication applications. Another example comes from semiconductor lasers. Recent work comparing Japanese and American laser diodes industries has concluded that the American system, with its vast VC-backed startup ecosystem, was able to experiment with more diverse applications in communications and medicine (Shimizu, 2019). Japanese laser diode research on the other hand was dominated by large incumbents, which focused on their core product markets in data recording and storage. Seminal Japanese corporate scientific discoveries in this field such as VCSEL (Vertical-Cavity Surface-Emitting Laser) were commercialized by American startup firms.

There are two types of explanations that account for startups commercializing more novel products than incumbents (Adner, 2002). The first focuses on "supply-side" factors that consists of the focal technology and firms' technical capabilities (Helfat and Lieberman, 2002). Certain technological "trajectories" may run their course and incumbent trajectories may yield more cost cutting process innovations than novel product innovations (Utterback and Abernathy, 1975). Startups may be technically more proficient at developing radically new innovations, especially if the technical knowledge required is cutting-edge and therefore harder to transfer. The hierarchical nature of larger incumbents could disincentivize the pursuit of diverse technical directions (Sah and Stiglitz, 1986; Agrawal et al., 2010) and managers may choose to focus on existing trajectories to incentivize internal researchers (Rotemberg and Saloner, 1994). Corporate researchers may also have a "Not-Invented-Here" bias against outside knowledge (Agrawal et al., 2010) which may be accentuated if the new

technological regime has "competence-destroying" aspects because they are too distant from the firm's existing capabilities (Tushman and Anderson, 1986; Henderson, 1993; Sørensen and Stuart, 2000). Incumbents may also fear cannibalizing existing market share due to product innovation (Arrow, 1962; Kamien and Schwartz, 1982). Indeed, incumbent inaction often accounts for the spawning and success of spinoff firms in the laser industry (Klepper, 2007; Klepper and Sleeper, 2005).

On the other hand, incumbents may be more efficient at integrating new knowledge because they have built up the absorptive capacity required to decode new scientific knowledge (Cohen and Levinthal, 1989). Some technologies may also disproportionately benefit from scale economies. Application of information technology to management, for example, has been hypothesized to cause "winner takes most" patterns that lead to higher firm concentration in industries (Autor et al., 2020). If this is the case, then incumbents may go further and "preempt" R&D in new technologies by outspending potential entrants (Gilbert, 2006) or acquiring early ones (Cunningham et al., 2021). At the core of supply-side explanations for startup superiority is an assumption that the new knowledge being commercialized "fits" better with technical capabilities that are path-dependent (i.e., are hard to build up rapidly). Recent work therefore has examined how firms in the pre-entry phase can acquire such capabilities rather than grow them (Moeen and Mitchell, 2020). These explanations would predict exit rates for startups to be lower if they commercialize new products, as both are correlated with a superior "fit" with the new technology compared to incumbents. By this reasoning, startups' entry rates conditional upon patenting should also be higher since their inventions are superior and patenting patterns should be different from incumbents. Since my theory relies on selection of startups into novel products, I also predict higher novelty and lower exit rates. However, I predict entry rates to be lower for startups precisely because of selection and do not require pre-entry activities (patenting patterns) to be different between

firms.

Demand-side explanations emphasize how firm and product characteristics interact with downstream market conditions (Schmookler, 1962). The "Schumpeterian" tradition in the economics of innovation has long studied how market structure affects incentives to invest in R&D (Schumpeter, 1950; Arrow, 1962; Gilbert and Newbery, 1982; Reinganum, 1983). While the earlier literature did not distinguish between types of innovations, recent work has shown that large firms generally specialize in incremental innovations while small firms in radical ones (Henderson, 1993; Link, 1982; Pavitt, 1987; Mansfield, 1981). A core assumption in the demand-side explanations is that capabilities are more responsive to incentives than supply-side explanations. For instance, Christensen (1993) argues that "disruptive" hard disk drive generations were readily pioneered by incumbents so far as their existing customers preferred them. Building on this, Adner (2002) argues that the types of market demand can affect product development strategies. However, these studies do not distinguish the invention stage before entry and the innovation stage after entry into the downstream market. An exception is Cohen and Klepper (1996) that shows how larger firms prefer process to product R&D because process innovations can be embodied in and spread across larger output.[7] This distinction is critical since downstream commercialization costs could incentivize startups to select into more ambitious projects which in turn increase their survivability upon entry. This mechanism does not require startups to be risk-loving; instead, higher entry costs disincentivizes entry of startups that settle for incremental (safer but less valuable) products. Observing firms at the invention stage allows me to count the "at bats"/"shots taken" (attempts at commercialization) and infer the existence of this selection.

[7]Note, though, that industry-level studies that actually observe the innovation output of product and process R&D are rare, as data discerning innovation types is difficult to find.

As well, the literature does not fully consider how the tradeoff between risk and reward for novel products can affect incumbents and startups differently. For instance, consumer preferences for a certain type of hard disk drive (Christensen, 1993) and the embodiment of process R&D in firm output are known properties for the firm making invention decisions (i.e., not uncertain before entry). However, innovations differ in their level of uncertainty — novel innovations may offer higher payoffs but may also fail more often (Fleming, 2001). Science can lower technical uncertainty at the invention stage (Fleming and Sorenson, 2004), but cannot reduce commercial uncertainty of how the new product would be received in the downstream market (Arora et al., 2019b). Yet, the management of such tradeoffs for new innovations is the central concern for practitioners commercializing new science (Lerner and Nanda, 2020; Howell, 2017; Gans et al., 2020). Therefore, the study of what kind of firm is more likely to exploit new science needs to separate the invention and innovation stages before and after entry such that the role of downstream uncertainty can be incorporated into the firms' entry decisions.

4.2.1 A simple model of science commercialization

I sketch a model to conceptualize the data generating process.[8] Firms commercialize new science in two steps. At the experimentation stage, firms invent new products and explore new markets at a cost (C) and a probability of success p. Innovation value x is distributed uniform with mean y and range b. b is higher for novel products (there is wider heterogeneity in their values). After the experiment is finished and x is realized, the firm makes the entry decision. Startups pay a larger cost of entry (Z) in the downstream market because they lack complementary assets. After entering, firms exit subject to a random shock ($\epsilon \sim Uniform[-a, a]$).

A firm experiments if the expected value of the new innovation exceeds the cost

[8]The proofs for the key findings are in Appendix C.1

$\mathbb{E}[X] - Z - C > 0$ (where $X = x + \mathbb{E}[\epsilon|\epsilon + x > 0]$). Likewise, the firm enters if the expected payoff is larger than the entry cost $p(\mathbb{E}[X|X > Z]Pr(X > Z) - Z) - C > 0$.

This implies that there is a minimum threshold value of the innovation (x^*) below which firms will not enter. Firms need higher innovation value to enter either if the downstream market is more unpredictable (i.e. the range of ϵ, (a) is higher) or if the entry cost (Z) is higher. Since the probability of entry is the share of innovations that are above this entry threshold, it is lower for firms that pay higher entry costs (i.e., startups). Other things being equal, all firms prefer more novelty (b) as it increases their expected payoff. Moreover, while startups earn less than incumbents for a given innovation because of higher entry costs, the marginal returns are higher for novel inventions. To see this, consider a twofold increase in b where $Z_{startup} = y$ and $Z_{incumbent} = y - b$ (for simplicity, also assume the downstream shock ϵ does not exist). Then, innovation payoffs increase by twice for startups ($= \frac{y+2b-Z_{startup}}{y+b-Z_{startup}}$), but only 1.5 times for incumbents ($= \frac{y+2b-Z_{incumbent}}{y+b-Z_{incumbent}}$). That is, novel inventions generate increased payoffs for incumbents, but not as large an increase as for startups. Therefore, startups are more likely to commercialize novel (high b) products. The higher thresholds also imply that the probability of exit is lower for startups upon entry, because exit is conditioned on the probability of entry.[9]

Table 4.1: Empirical Predictions

Result 1	startups enter the product market less often than incumbents
Result 2	startups are more likely to commercialize novel products than incumbents
Result 3	startups exit the product market less often than incumbents

In other words, startups choose more novel innovations because their payoff threshold for entering is higher. Startups have a disadvantage in existing markets

[9]The exit result requires random shocks in the downstream market - if all eventualities in the innovation phase are foreseen during experimentation, the entry threshold would simply be the entry cost (Z) and no firm would exit after entering.

and incremental innovations (low b options), since they lack complementary assets such as manufacturing, sales and marketing (Teece, 1986; Chatterji, 2009). However, this disadvantage is less prominent for novel markets and innovations where incumbents must also be "newcomers". Having an absolute disadvantage in incremental innovations but not in novel ones implies that startups have a comparative advantage in the latter. Since startups have an absolute disadvantage vis-a-vis incumbents, this would increase their entry threshold and incentivize pursuing novel and uncertain innovations, where they have a comparative advantage in. That is, startups that enter the market are those that have traded off certainty of success with higher payoffs to exceed the entry threshold. Conversely, incumbents with lower entry thresholds can afford to trade off lower expected payoffs with more certainty (i.e., enter into incremental innovations). This formalizes the speculation by Griliches (1990) that the apparently higher R&D productivity of smaller firms in large scale patent data may be attributable to selection. I test this hypothesis in Section 4.5 using entry patterns of laser firms whose products are based on Soviet science.

4.3 Empirical Context: Laser Science Across the Iron Curtain

A simple comparison between the types of innovation done by startups and incumbents may be confounded by either type of firms having preferential access to the source technology. Startups formed by university scientists, for instance, could have better access to tacit knowledge that is hard to emulate for incumbents.[10] Conversely, incumbent firms with prior contracts that demand substantial investments in cutting-edge research (e.g. early defense contracts for the semiconductor and computer industry (Mowery and Rosenberg, 1998)) may tap into internal scientists to

[10]Even within the same firm, cutting edge research is often hard to communicate without direct human contact. A prime example is the Scanning Tunneling Microscope which was invented at IBM's Zurich laboratory - multiple trips needed to be arranged between Yorktown Heights and Zurich for American teams to replicate the microscope (Mody, 2011, pp.69-70).

innovate new products. This calls for variation in new knowledge that is not directly linked to startup or incumbent characteristics.

Therefore, I use the fall of the Berlin Wall as a source of unexpected variation in scientific knowledge that does not systematically favor startups or incumbents. The laser sector is unique in that scientific papers, patents, and products can be directly linked through the gain media and optical pumping method used (for a detailed discussion, please see Section 4.4). Moreover, laser science is an area where the Soviet Union excelled in. This allows me to use the fall of the Soviet Union as an exogenous shock to knowledge hitherto unknown (or costly to acquire) in the West (Borjas and Doran, 2012; Ganguli, 2015; Teodoridis et al., 2019). Results in Section 4.5.2 shows that startups and incumbents did not significantly differ in their propensity to absorb this knowledge, adding confidence that there was a relatively level playing field between firms for areas that gained access to Soviet knowledge. Instruments at the firm level such as R&D tax credits and defense contract reassignments have been used as a type of exogenous variation in knowledge inputs (Moretti et al., 2019; Bloom et al., 2013; Arora et al., 2021b). However, direct linkages between products and scientific fields are still rare, with notable exceptions in the life sciences where progress in commercialization is easier to track through regulatory disclosure (Moeen and Agarwal, 2017; Krieger et al., 2018; Kao, 2019; Jayaraj and Gittelman, 2018).[11]

Lasers build on the principles of stimulated emission that were first formalized by Albert Einstein in 1917. In its simplest form, stimulated emission refers to the phenomenon whereby atoms emit electromagnetic waves as they return to their "normal" lower energy state from an excited, higher energy state. The medium containing the atoms being excited is called the "lasing medium" or "gain medium", while

[11]Recent work has linked scientific articles to patents through non-patent citations (Marx and Fuegi, 2020; Arora et al., 2021b), but patents should more appropriately considered as "inventions" that feed into the production function of a firm that must produce "innovations" (products sold on the market) to qualify as a commercializer of scientific knowledge.

the energy source is called an amplifier: hence, the acronym, Light Amplification by Stimulated Emission of Radiation (LASER). Lasing media can range from solid crystals, liquid dye, and various gases, while amplifiers can range from flash lamps to electric currents (for diode lasers). Laser research focuses on sustaining the "population inversion" of atoms between these disparate energy states. The end product is a collimated (straight) beam of light with high monochromaticity (single color) and narrow spectral linewidth. Further engineering developments allowed for modulating the pulse length and power of these beams (e.g. Q-switching and mode-locking), while research has continued on expanding the types of media that are amenable to population inversion. Both American and Soviet scientists researched lasers and arrived at its invention in roughly the same period in the early 1960s. However, differences in scientific organization across the iron curtain resulted in subsequent researchers to specialize in different subfields of laser science.

In the United States, laser research was inaugurated when Charles Townes and Arthur Schawlow achieved stimulated emission at the microwave wavelength using ammonia gas in 1953. Based on follow-on research, the authors theorized in 1958 a method to achieve the same feat at the visible range (Schawlow and Townes, 1958)[12], and two years later TRG's Ted Maiman had invented a working laser by using a Ruby crystal pumped by a flashlamp. Nikolai Basov and Alexander Prokhorov in the USSR had also arrived at this solution independently, with theoretical sketches occurring as early as 1952. The pair were co-laureates of the Nobel Prize in physics in 1964 together with Townes and Schawlow for this research. Subsequently, Prokhorov and Basov directed laser research programs around the USSR with extensive state support, and Soviet scientists were credited with the first working excimer laser in 1971 (Basov, Danilychev, Popov) as well as the synthesis of quantum dots in 1981 (Alexey Ekimov). Areas of (comparative) Soviet excellence in laser research gradually

[12]In the paper, they refer to the laser as an "optical" maser.

emerged in rare-earth element doped glasses and crystals, as well as semiconductor lasers (Bagayev et al., 2005; Prokhorov and Shcherbakov, 1991) separately from Western developments. Hence, Soviet laser research was overall not far behind its American counterpart, while they had parity in select sectors such as semiconductor lasers and superiority in others such as rare-earth doped solid state lasers.[13]

The divergence in research focus was due to historical circumstances owing to the organization of laser research in the two systems. In contrast to the American system that relied on a combination of government research centers, corporate R&D labs and startups, laser research in the Soviet Union was concentrated in a handful of state-sponsored institutes. Under a centralized system, opening radically new areas of research may have been harder, but funding for existing streams of research may have been steadier. For instance, both the US and the USSR had initiated glass laser research, but the Soviet Union was the first to develop glasses capable of maintaining stable temperatures necessary for high power, pulsed lasing conditions ("athermalization").

In spite of these achievements, Soviet laser science could only be incompletely communicated to the West during the Cold War. Translations of Soviet journals were often sparse and delayed, and translated versions were expensive[14] and "left many questions unanswered" (Hecht, 1999, p.93). Laser inventions deemed strategic in nature were intentionally kept secret by the Soviet state. For instance, out of Alexander Prokhorov's eleven published inventions, seven were censored until the USSR's collapse (Martens, 2020).

With the collapse of the Soviet Union, state censorship of these materials was relaxed, while the migration of scientists became freer. Not only did the United

[13]Please see Appendix C.2 for a more detailed historical overview of Soviet laser research.

[14]Up to twenty-eight times the price of their Russian language originals according to some accounts (Hollings, 2016).

States relax entry requirements for Eastern bloc scientists through the Soviet Scientist Immigration Act of 1992, but the working conditions for scientists were also becoming unfavorable in Russia, with high inflation and loss of state-sponsored jobs (Mirzabekov, 1993). This led to the emigration of Soviet scientists who would further reduce the tacit information barriers that existed during the Cold War (Ganguli, 2015). Some laser scientists, such as Dmitri Basov (Nikolai Basov's son), took up jobs at academia (Columbia), while others such as Dmitri Garbuzov (Alferov's colleague), founded startups (Princeton Lightwave).

Many prominent laser firms count Soviet-trained inventors in their midst during the 1990s. For instance, well over 40% of patents at Cymer[15] filed from 1989 to 2015 were invented by two Russian scientists (Alexander I. Ershov, Igor V. Fomenkov). Moreover, lasers in areas of aforementioned Soviet superiority began to be commercialized in products that were fundamentally different: the Soviet expertise in rare-earth elements, for instance, was realized when Valentin Gapontsev introduced the world's first commercial fiber laser above 10W in 1996.[16] These lasers use glass fibers doped with rare earth elements such as Erbium and Ytterbium as the gain media, and are capable of operating at higher power and efficiency compared to conventional lasers.

The parallel but divergent development of laser physics across the Iron Curtain and the subsequent fall of the Soviet Union offers an exogenous shock to the cost of using superior scientific knowledge for select subfields in lasers for American firms. Cold War era information barriers ensure that neither startups and incumbents had preferential access to this research as the Berlin Wall fell. Section 4.5.2 tests the

[15]Cymer was the dominant Deep Ultraviolet (DUV) excimer laser firm that enabled the introduction of 256MB DRAM chip fabrication process in 1998 as semiconductor firms switched from xenon lamps to DUV lasers for photolithography.

[16]Gapontsev received his PhD from the Moscow Institute of Physics and Technology and founded IPG Photonics in 1990.

validity of this assumption by comparing patenting behavior of the two types of firms after 1990.

4.3.1 Reallocation of Cold War Defense Spending

A potential concern with my setting is that the end of the Cold War also coincided with a significant reallocation of government science spending which may simultaneously encourage novelty product introductions and the entry of startups. Increase in public funding for R&D in a scientific discipline has been shown to boost patenting and innovation in related areas (Lichtenberg, 1987; Azoulay et al., 2018; Fleming et al., 2019; Moretti et al., 2019). The laser industry was also a beneficiary of public funding support.[17] The years surrounding the fall of the Berlin Wall also saw a significant reallocation in government spending in lasers. The Reagan administration's Strategic Defense Initiative (SDI), for example, began in 1983 and most prominently funded research into Directed Energy Weapons (laser beams with high intensity to intercept incoming ballistic missiles). With the reduction of nuclear weapons stockpiles after the end of the Cold War, political support for SDI waned and the project was terminated in 1993.[18] Given that this reallocation of public laser funding occurs at roughly the same time when Soviet science is made public to Western scientists, any empirical design comparing difference in laser firm characteristics before and after the Cold War may identify the effect of U.S. government funding rather than of the availability of new knowledge from the Eastern bloc.

[17]The first (Ruby) laser, for example, was invented by a scientist at an aerospace & defense contractor (Ted Maiman at the Hughes Aircraft Company). Gordon Gould, the author of the first "Laser Memo", also obtained a position at TRG (another recipient of ARPA funding) after finishing his doctoral studies at Columbia University. Indeed, of the three potential areas of application (measurement, communication and materials processing), measurement was commercialized fastest due to funding by the National Bureau of Standards and defense funding during the Vietnam war (for details, please see Appendix C.3.2).

[18]The agency in charge of SDI was renamed into the Ballistic Missile Defense Organization in 1993 and research into Directed Energy Weapons was cut. Appendix figure C.1 shows that the end of the Cold War was associated with a precipitous decline in defense contracts for lasers.

Government contracts may bias estimates on who exploits new science from the demand side. On the one hand, such contracts may favor large incumbent firms that have the manufacturing capabilities to supply vast procurement requirements. On the other hand, set-aside provisions that explicitly give priority consideration to small businesses would increase startup entry in funded areas.[19] From the supply side, government funding reallocations may give a head-start to contracting firms by either directly funding research or incentivizing research meant to satisfy performance requirements (Belenzon and Cioaca, 2021). If contract reallocations after the fall of the Berlin Wall were primarily toward lasers in which the Soviets concentrated in, then this undermines the assumption that such lasers opened a level playing field. I mitigate this concern through two robustness checks. First, I exclude the lasers that are related to the SDI program and replicate my results. Second, I test whether the coefficient estimates for product novelty and startup status are larger in magnitude for American cities where Soviet scientists and engineers immigrated more into after the fall of the Berlin Wall. Prior research from Ganguli (2015) show that the immigration of Soviet scientists increased awareness (measured by subsequent backward publication citations) of Soviet-era research by American scientists. This migration-based geographical variation should be orthogonal to government procurement funding.

4.4 Data

I combine product buyer guides, patents, and scientific articles to trace innovations (products) to their underlying inventions (patents) and back to the background science (physics papers). My sample ranges between 1980 and 2010 to exploit the end of the Cold War and observe the years before and after this event. To my knowledge,

[19]Small business set-asides were first instituted in the Small Business Act of 1958 and subsequent amendments have increased the annual minimum small business contracting goals to 23%.

Table 4.2: Summary of Variables

Stage	Variable Name	Proxy for	Definition	Unit of Observations
Pre-Invention	*SOVIET PAPER SHARE*	new scientific knowledge	Share of laser papers published in Soviet journals (1980-90)	optical gain medium
Invention	*SMALL*	firm size	1 if assignee small entity classified under section 41 of U.S. Code, 0 otherwise	assignee
Invention	*COMPUSTAT*	firm size	1 if assignee matched to DISCERN database of public U.S. firms, 0 otherwise	assignee
Invention	*NOVELTY*	invention novelty	1-textual similarity scores for patents	patent
	ENTRY	firm entry into product market	Assignee name matches seller in Laser Focus Buyer's Guide	assignee
Innovation	*STARTUP*	firm size	1 if firm age under five years, 0 otherwise	firm-year
Innovation	*FIRST*	innovation novelty	1 if first product in construction-beam type-gain medium combination, 0 otherwise	product
Innovation	$\|Z\ SCORE\|$	innovation novelty	Absolute Z-scores for divergence from mean for wavelength, power, repetition rate, pulselength, repetition rate, and beam diamaeter	product-year
Innovation	*SALES JUMP*	market novelty	Dummy for whether there is a 200% sales jump in a new application (out of 13)	product
Innovation	*SALES HHI*	market novelty	Herfindahl index of sales distribution over 13 applications	product
	EXIT	firm exit from product market	3, 5, 10 year survival rates ("death" occurs when firm-product is discontinued in the Buyer's Guide)	firm-product

this is the first dataset that links basic scientific research to product innovations for firms outside of the life sciences (Krieger et al., 2018; Kao, 2019; Jayaraj and Gittelman, 2018).

4.4.1 Science

To identify basic laser research fields where the Soviet Union had a comparative advantage in, I rely on the *Handbook of Laser* (Weber, 2000), an authoritative text in the discipline that documents the first time lasing phenomena are reported for 18,313 lasers by their optical gain medium. These lasers are linked to one or more of 4,046 scientific papers that are credited with each discovery.[20]

[20]Most lasers are linked to fewer than three papers.

Soviet Paper Share — Based on expert consultation and historical readings, I manually identify fourteen Soviet journals in which lasing is reported in Weber (2000).[21] Soviet publications account for around 24% (262) of the 1,090 papers published between 1980 and 1989. Soviet publications tend to be concentrated in paramagnetic solid state lasers (in the early 1980s, over half of all solid state laser publications are found in Soviet journals), while less concentrated in gas and dye lasers relative to the rest of the world. Table 4.3 presents a detailed breakdown by the type of gain medium studied and finds that rare earths elements such as Ytterbium, Holmium and Erbium (used in solid state and fiber lasers) have the highest Soviet paper shares, while Hydrogen Fluoride and Carbon Monoxide (used in gas lasers) have no Soviet publications in the 1980s.

[21]The journals are: *Doklady Akademii Nauk, Soviet Journal of Quantum Electronics, Physics Dokladi, Bulletin of the Russian Academy of Sciences: Physics, Journal of Russian Laser Research, Uspekhi fiziologicheskikh nauk, Soviet physics. Solid state, Soviet physics. Crystallography, Journal of Experimental and Theoretical Physics, Soviet physics, JETP, Soviet physics. Technical physics, Soviet physics. Semiconductors*

Table 4.3: Soviet Paper Share by Gain Medium (1980-90)

Gain Medium	Soviet Paper Share (1980-1990)	Gain Medium	Soviet Paper Share (1980-1990)
Yb	0.85	HeCd	0.06
Ho	0.45	XeHe	0.04
Er	0.44	HeXe	0.04
Mn	0.43	Xe	0.04
Tm	0.41	metalvapor	0.04
Nd	0.39	excimer	0.03
Ti	0.33	Cu	0.03
Ba	0.30	CH3OH	0.02
color_center	0.28	Kr	0.02
Ce	0.25	Ar/Kr	0.02
Pb	0.23	HeNe	0.01
semiconductor	0.20	Ar	0.01
KrCl	0.17	N2	0.00
Far_IR	0.15	KrF	0.00
CO2	0.15	NH3	0.00
Cr	0.15	CO	0.00
F	0.11	CH3F	0.00
XeF2	0.11	ArF	0.00
HeSe	0.10	XeCl2 / Xe2Cl	0.00
glass	0.10	H2O	0.00
HeAu	0.08	HF/FH	0.00
Au	0.07	XeBr2	0.00
I	0.06	GaAlAs	0.00
HCN	0.06	HF	0.00
HeAg	0.06	Co	0.00

Notes: Unit of analysis is the optical gain medium. Sample consists of publications found in the Handbook of Lasers (Weber, 2000). "Soviet Paper Share" divides for each gain medium the number of papers published in Soviet journals by all papers in that gain medium during the 1980s. Observations are downward sorted by Soviet paper share.

4.4.2 Invention

I rely on patent data to identify pre-entry inventive activity and calculate entry probabilities for patentees. I source patent data from Google Patents (IFI Claims Services) and USPTO Patentsview for U.S. utility patents with grant years between

1980 and 2010.

Reliance on (Soviet) Science — I measure reliance on science through front page NPL and in-text citations made to the scientific literature from patents. My source data is Marx and Fuegi (2020), who match US patents to scientific publications in Microsoft Academic Graph (MAG) to identify pairs of citing patents and cited scientific publications. I identify around half (11,998) of the 24,903 laser patents to be citing at least one scientific article in MAG. Of these, 245 cite papers published in the Soviet Union (please see the subsequent section for details on identifying Soviet laser publications). In addition to this "direct" link, inventors may also become aware of Soviet scientific knowledge indirectly through American papers that cite such articles. Therefore, I follow Ahmadpoor and Jones (2017) and calculate "degrees of connection" between patents and Soviet scientific publications, where a patent citing a Soviet paper receives a degree of zero (and a patent which cites a paper which in turn cites a Soviet scientific paper receives a degree of one). I calculate up to three degrees of connections, where 1,388 patents cite Soviet science at the first degree, 4,733 at the second and 8,761 at the third degree.[22]

[22]I stop at the third degree as it covers around 65% of laser patents that cite any scientific article.

Table 4.4: Descriptive Statistics for Laser Patents

	Count	Mean	Median	Std Dev	Min	Max
Patent Publication Year	24903	2001.93	2003	9.12	1980	2016
Forward Patent Cites	24903	0.74	0.38	2.02	0.00	210.96
Soviet Paper Share (Patent)	17022	0.14	0.07	0.16	0.00	0.85
Patent Cites to Science	24903	4.51	1.00	25.06	0.00	763.00
Patent Similarity	15438	0.24	0.23	0.10	0.02	1.00
Cites to Soviet Science (D=0)	24903	0.01	0.00	0.10	0.00	1.00
Cites to Soviet Science (D=1)	24903	0.06	0.00	0.23	0.00	1.00
Cites to Soviet Science (D=2)	24903	0.19	0.00	0.39	0.00	1.00
Cites to Soviet Science (D=3)	24903	0.35	0.00	0.48	0.00	1.00

Notes: Sample consists of patents whose 4-digit CPC classification includes "H01S". *Forward Patent Cites* divides the total number of forward patent citations received until 2016 and divides them by the average forward cites received by patents granted in the same year as the focal patent. *Soviet Paper Share (Patent)* refers to the share of scientific articles published in Soviet journals during the 1980s, matched by the IUPAC International Chemical Identifier (InChI). *Patent Cites to Science* counts the number of citations made to scientific journals from the focal patent (Marx and Fuegi, 2020).

Linkage to Gain Media (InChI) — Patents are mapped to Soviet paper shares via the gain media listed in them. I source data on chemical compounds mentioned in the body of the patent text from Google patents. This contains the full chemical formula as well as the IUPAC International Chemical Identifier (InChI). I then manually search for the InChI codes for the 50 chemical compounds found in the Laser Handbook from PubChem, a public search engine maintained by the National Library of Medicine.[23] I then restrict the patent data to only include laser patents (CPC: H01S)[24] and link the Soviet publication share during the 1980s to patents using their InChI code. Where there are more than one InChI codes linked, I keep the chemical compound with the highest Soviet paper share for that patent.

[23]https://pubchem.ncbi.nlm.nih.gov

[24]This is crucial because patents from other classes (especially in the biochemical and semiconductor sector tend to make heavy use of InChI codes.

For example, US9147992B2 is a fiber laser systems patent by Coherent that uses certain rare earth elements such as Erbium, Ytterbium and Thulium.[25] Google patents extracts the InChI codes for each (InChI=1S/Er, InChI=1S/Yb, InChI=1S/Tm), which I link back to Soviet publication share data in the Laser Handbook (44% for Er, 85% for Yb, 41% for Tm). The Soviet publication share score for the patent therefore is 85%.

Invention novelty — I use *patent textual similarity* to prior patents to measure invention novelty. Building on Arora and Cohen (2018); Arora et al. (2018a), for each focal invention, I calculate its textual similarity score for all previous patents (all USPTO patents with an earlier priority date than the focal invention). I normalize the proximity scores vector of the top 100 closest citation pairs for each focal patent by dividing each score by the corresponding maximum pairwise textual score for the focal patent. I average the standardized scores to derive a single textual proximity score for each focal invention. On average, an invention has an aggregate textual proximity score of 0.24 with a standard deviation of 0.10 and a median of 0.23. Subtracting this score from one yields the invention novelty score. A lower score indicates less novelty (greater textual overlap with existing inventions), while a higher score indicates more novelty.

Invention quality — I use three measures of invention quality. First, I use data from EPO Patstat to count the number of *forward patent citations* a patent has received and normalize this by the average number of citations received by all patents in the focal patent's publication year. The second measure is a dummy indicating whether

[25] "pulsed laser output by propagating the pulsed laser output through a non-silica-based gain medium that has been doped with a concentration of a rare earth ion, such as **Erbium, Ytterbium, and/or Thulium**, wherein the concentration of the rare earth ion within the gain medium is approximately greater than one percent by weight." (emphases added by author, full patent at https://patents.google.com/patent/US9147992B2/en?oq=US9147992)

a patent is triadic patent registered in the three largest patent jurisdictions: U.S., European and Japanese patent offices (Dernis and Khan, 2004). Patenting in all three offices is consistent with the invention being valued highly by the assignee. I also count the number of claims and the length of the first claim as additional measures of patent quality.

Startup — I identify startups based on the declared size of the patent's assignee in the USPTO maintenance fee payment records. Firms with less than 500 employees are subject to 50% lower filing and maintenance fees per USPTO regulations. I also identify larger incumbent patentees matched to the DISCERN database (Arora et al., 2021b), which matches listed U.S. firm names from S&P Compustat to USPTO patents.

4.4.3 Entry

If a patent assignee in the laser (H01S) CPC is matched to the *Laser Focus* Buyer's Guide, I code it as an entrant into the product market.[26] I standardize the manufacturer name found in the Buyer's Guides, removing company suffixes (LTD, INC, PLC) and names of business units. In case a firm becomes a division within another firm due to M&As, I reassign products of the firm to the parents post acquisition. I identify abbreviations such as SDL (Spectra Diode Labs), SAT (Société Anonyme de Télécommunications) and NEC (Nippon Electric Company). I also identify name variants such as Bell Labs and Watson Research Center for AT&T and IBM. I then manually match these names to the 3,450 assignees with at least one patent in the laser CPC (H01S) during our sample period.[27] Around 43% of firms patent at least once during our sample period.

[26] Please see the next section for details on this data source

[27] This is necessary given the occurrence of similar names in the laser industry (e.g. names such as Spectra-Physics, Spectral Systems, Spectro Laser System would yield very similar scores using conventional string distance metrics.

4.4.4 Products

I use a leading industry publication, *Laser Focus*, to collect data on firms and the products they field in the American laser market between 1980 and 2010.[28] I link the products to Soviet science using their gain media, calculate their technical novelty based on product attributes, and measure commercial novelty by counting how many new applications the lasers were sold to.

Table 4.5: Descriptive Statistics (Firm-Year)

	Count	Mean	Median	Std Dev	Min	Max
Year	2569	1994.53	1995	8.82	1980	2010
Founding Year	2115	1974.7	1978	22.36	1847	2006
Startup	2569	0.14	0	0.35	0	1
Number of Employees	1631	126.36	30	556.98	1	18000
Soviet Paper Share	2510	0.2	0.2	0.14	0	0.85
Number of Laser Products	2569	29.53	9	86.87	1	1633
Number of Laser Types	2569	2.66	2	2.52	1	19
Number of Patents	2569	16.67	0	96.41	0	1911

Notes: Unit of analysis the firm-year. Sample consists of firms in *Laser Focus* Buyer's Guide that produce at least one laser. *Founding Year* refers to the year the laser firm was established according to the vendor directory. Firm age is calculated by subtracting the founding year from the panel year, and *Startup* is a dummy equal to one if firm age is less than five years. *Soviet Paper Share* refers to the share of Soviet publications linked to the laser type that a firm produces. *Number of Laser Products* counts the number of products sold each year. *Number of Laser Types* counts the number of construction-beam type-gain media combinations sold by the focal firm and year. *Number of Patents* refers to the number of patents granted to assignees matched to the firm names.

Product Novelty — The preferred measure of product novelty measures whether a product type has never been introduced before. I define product type by the combination of construction (e.g. Tunable Solid State), beam type (e.g., pulsed) and gain medium (e.g. Neodymium). There are seven methods of construction, two beam types, and fifty gain media. 123 product types out of these 700 ($= 7 \times 2 \times 50$)

[28]For previous papers using this publication, see Klepper and Sleeper (2005); Bhaskarabhatla and Klepper (2014); Sleeper (1999).

possible combinations have been commercialized by 9,180 products in my sample. Most product types are fielded by multiple products and firms. For example, there are 45 products classified as Tunable Solid State, Pulsed lasers using Neodymium (Nd) as their gain medium in the buyer's guide (please see Appendix Table C.2 for details on all product types.). The first recorded appearance of this product type is in 1994, when Continuum Inc sold 5 products of this type (the model numbers are HRL-1, HRL-100, HRL-1002, HRL-50C, and Mirage). Based on this, I code Continuum's five lasers in 1994 as a "first" product in their type (while the 40 subsequent lasers in that type receive a value of zero for this dummy variable).

I also measure the novelty of a laser sold on the product specification tables by calculating the divergence of its lasing characteristics from the average value for the lasers in the same construction and year. The former (product-type based) measure is concerned with how the laser is built, while these measure what the laser is capable of. Two different types of lasers can have the same wavelength or power, for instance. The novelty score for laser i with wavelength $wave_i$, produced in year t and classified as construction-beam type combination k is:

$$|Z_Wave_{ikt}| = |\frac{wave_{ikt} - \overline{wave_{kt}}}{\overline{wave_{kt}}}| \tag{4.1}$$

The same calculation is repeated for laser power, beam accuracy, and spectral linewidth. For example, the average power of a pulsed gas laser in 1996 was 2,231 Watts, but Rofin Sinar's UR150 Carbon Dioxide laser is registered at 15,000 Watts. Therefore, the absolute Z score for its wavelength is $|\frac{15000-2231}{2231}| = 5.7$. For an opposite example of a laser close to the mean, consider Coherent's 899-21 dye laser in 1996, which has an output power of 0.860 Watts. Given that continuous wave dye lasers' average power is .844 Watts that year, the absolute Z score for this laser is low ($|\frac{(0.860-0.844)}{0.844}| = 0.02$). Note that attributes such as wavelength are not linearly correlated with quality (i.e., a laser is not necessarily better or harder to construct because it has higher wave-

length). This motivates using absolute values of the Z-scores to measure only the deviation from the mean as a measure of novelty.

Products Linkages to Science — The *Laser Focus* product specification tables contain the chemical formula of the gain medium, which I use to link Soviet paper shares to products. For instance, the Soviet paper share in Helium Neon (HeNe) lasers is 1.3%, while that for lasers using Ytterbium (found in various types of solid state (e.g. Yb:YAG) and fiber lasers) is 85%. There are 50 gain media that are linked between the Handbook and the LFWBG specification table.[29]

Product Submarkets — I link lasers to downstream markets based on the annual sales data for thirteen applications from the *Laser Focus* Annual Market Reviews, collected between 1986 and 2005.[30] These sales figures are broken down by 26 laser types which loosely correspond to optical gain media.[31]I map laser products to markets based on the most granular possible level of information consisting of construction, gain media, wavelength, and power (i.e., match on gain media where market data is available; power and construction where gain medium is not available and so forth).

[29]To simplify the correspondence, for solid state lasers, I map based on dopants and do not use information on crystals. For instance, an Nd:YAG laser in the specification table would be mapped based on the Neodymium (Nd) dopant as the gain medium rather than the Yttrium Aluminum Garnet (YAG) crystal. This allows for a more accurate linkage between science and products because there is typically more variation in the types of dopants used than the garnets (YAG crystals may be doped with Erbium, Neodymium, Holmium, Ytterbium). Moreover, garnets themselves tend to be compounds (e.g. YAG, which is a Yttrium Aluminum Garnet; YLF, which is a Yttrium Lithium Fluoride compound), making it difficult to do one-to-one linkages at the atomic level. Also, while most gain media are used for one type of construction (e.g. CO2 for gas lasers), others may be used for multiple types (e.g. Erbium used in solid state lasers as well as fiber lasers).

[30]The thirteen downstream applications are: Material processing, Medical therapeutics, Instrumentation, Research, Telecommunications, Optical storage, Entertainment, Image recording, Inspection, measurement, and control, Barcode scanning, Sensing, Optical Pumping, and Other.

[31]Some gain media such as gallium arsenide used in semiconductor lasers have large market such that the *Laser Focus* editors decided to collect detailed sales figures broken down by lasing wavelength. By contrast, gas ion lasers comprise a variety of gain media such as Xenon Fluoride and Gold that are lumped together for the purposes of the market review (note though that more prominent gas gain media such as CO2, Helium Cadmium and Helium Neon are given separate categories).

To identify new applications, I count for each year and each of the 26 laser types the number of application areas that exhibit a sales jump in excess of 200% in the following year. To measure the downstream application generality related to each laser type, I calculate the Herfindahl-Hirschman Index for each laser across applications. For laser type k sold in downstream segment j:

$$HHI_k = \sum_j s_{kj}^2 \tag{4.2}$$

where s_{kj} is the share of laser k's sales that are attributable to market segment j. In my sample, the laser type with the highest average HHI (least general) is Flowing CO2 lasers, where 99.9% of all sales are made to materials processing (i.e., welding & machining). The lasers with the lowest average HHI are high-powered (> 1W) near infrared (750-980nm) diode lasers, which have sales in all categories except for barcode scanning, entertainment and optical storage. This allows me to identify flowing CO2 lasers as the least general and the high-powered near-infrared diode lasers as the most general lasers.

Startup — I collect information on firm founding year from the *Laser Focus* manufacturer's directory. I code a firm as a startup if its age in the focal year is five years or less. Age is calculated by subtracting the firm's founding year from the focal year (average age of a firm in my sample is 17 years).

4.4.5 Exit

Survival in the product market is defined for firm and product type (product types defined above as a combination of construction type, beam characteristic & gain medium). The "3-year survival" dummy equals one if the firm-product type continues to appear after three years after its first appearance in the Buyer's Guide. For startups, I ensure "successful" exits due to acquisition by another firm are not coded

as exits.[32] To identify such activity, I collect the industry news reports section (the Laser Industry Report, Optics Industry Report, Imaging and Detector Industry Report, Fiberoptics Industry Report, and the Postdeadline News Reports) as well as the business review sections for all monthly issues of *Laser Focus* and *Laser Focus World* between 1980 and 2010 to find instances of M&A, name changes and bankruptcies.[33] I collect the new name in the case of a name change or an acquisition. If a startup disappears due to an acquisition and the acquiring firm still fields products in the area, the startup inherits the survival years of the acquiror (i.e., it does not exit on the year of the acquisition). I also search the web where the industry reports do not mention any activity.

4.5 Results

A key challenge in identifying the role of entry costs to explain differences in novel product introduction rates between startups and incumbents is that some firms may have better access to the underlying science. Therefore, I leverage a geopolitical event that is exogenous to firm types to test my theory. The unexpected collapse of the Soviet Union began with the fall of the Berlin Wall in November 1989 and liberalized access to research that was produced in the USSR for Western firms. As explained in Section 4.3, Soviet laser science not only was high caliber but concentrated often in different subfields. This implies that the fall of the Soviet Union reduced the costs of accessing novel science hitherto unknown to both startup and incumbent laser firms in the West. In the following sections, I first establish a stylized fact that the abrupt availability of new laser science from the Soviet Union led to new

[32] As it is often the case in other technology-intensive industries, laser startup founders often aspire to be acquired by large integrated firms.

[33] In cases where only announcements of intent are reported I still code it as an acquisition unless there is a subsequent report reporting a withdrawal of offer or breakdown in negotiations. Distribution and marketing agreements are also excluded.

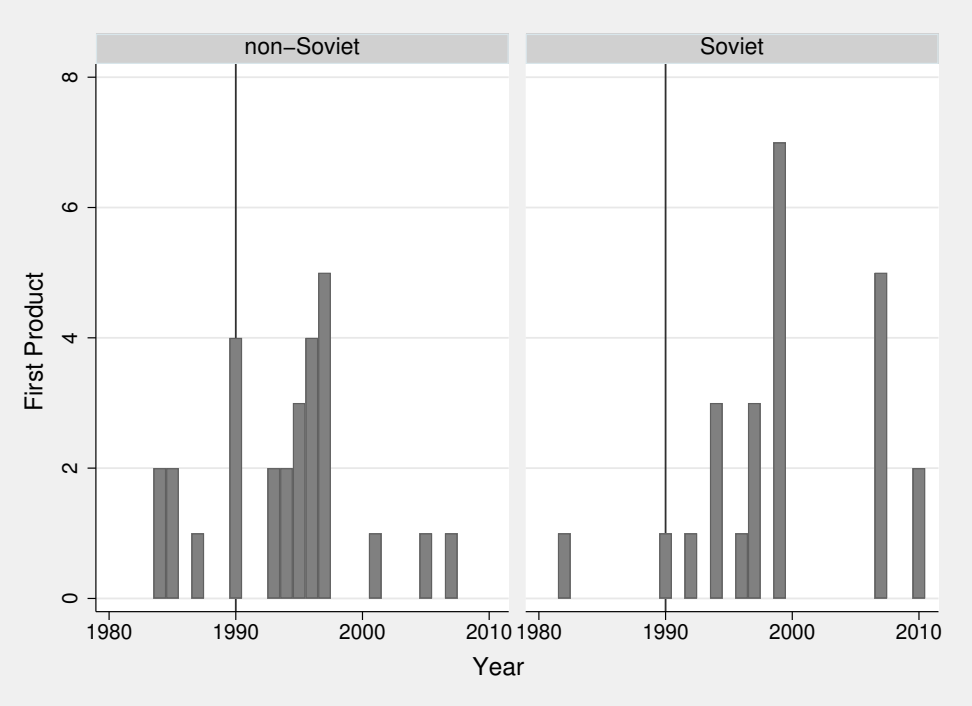


FIGURE 4.1: Annual Product Type Introductions, by Soviet Paper Share[34]

inventions (patents) and innovations (products). I then verify that startups were not more likely than incumbents to invent Soviet-type laser patents. To test Result 1 in Table 4.1, I test whether startup patentees were more likely to enter into the downstream product market. To test Result 2, I test whether products introduced by startups were technically novel and whether they were sold in new markets. Finally, I examine whether exit rates of startups are lower than incumbents per Result 3. To mitigate concerns the U.S. defense spending reallocations are driving the results, I replicate in Appendix C.4.2 the main results without lasers affected by the SDI and in Appendix C.4.3 test whether the effects are stronger for cities with more Soviet scientist migrants.

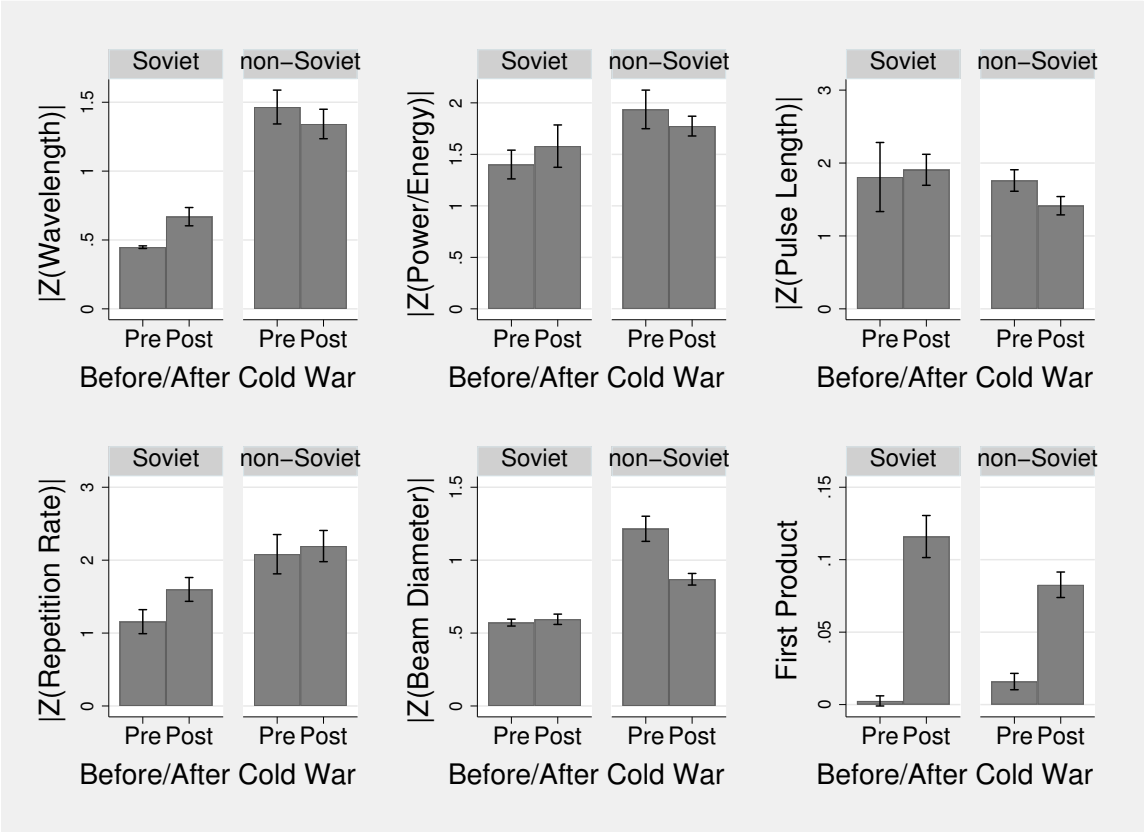


FIGURE 4.2: Product Novelty Before & After 1990, by Soviet Paper Share[35]

4.5.1 Stylized Facts on Soviet Science, Inventions and Innovations

Figure 4.1 plots the number of "first products" introduced each year by whether the laser is linked to a gain medium where the Soviet paper share during the 1980s is the in the fourth quartile (above 39%). "First products" are defined as the first product of a laser type (i.e., no laser of the same type has appeared before the focal year).

[34]This bar graph plots the number of new laser product types sold in the *Laser Focus* Buyer's Guide between 1981 and 2010. New laser product types are defined as the first construction-beam type-gain medium combination found in the Buyer's Guide.

[35]Sample consists of laser products sold in the *Laser Focus* Buyer's Guide between 1981 and 2010. "Pre" and "Post" periods each refer to 1981-1990 and 1991-2010. The "Soviet" and "non-Soviet" rows split the products by whether their gain media (linked to shares of Soviet publications in the Handbook of Lasers) had Soviet paper shares above or below mean shares. "—Z(Wavelength)—" refers to the absolute values of the divergence from the mean for laser product wavelength. The rest of the Z-scores are defined identically for the respective product attributes. First Product" is a dummy equal to one if the product is a first product type.

"non-Soviet" laser products continue to be introduced after the end of the Cold War, but rare after 1997 (when mainstream DUV excimer lasers were introduced). By contrast, there are more "Soviet-type" inventions introduced and entries continue into the 2000s when various improvements on fiber lasers are introduced. This is consistent with Soviet laser physics being the basis for novel innovations. Figure 4.2 plots the pre-post contrasts for Soviet vs non-Soviet lasers for this first product dummy together with the absolute Z scores for laser wavelength, power, pulselength, repetition rate and beam diameter. "Soviet-type" lasers are more likely to be a first product and diverge farther from the mean after the end of the Cold War in all five product characteristics, though the pre-post difference is statistically significant only for wavelength and the first product dummy.

I run a difference-in-difference specification for product i sold in year t,

$$Novelty_{it} = \beta_0 + \beta_1 Soviet_Share_i \times Post_1990_t + \beta_2 Soviet_Share_i + \boldsymbol{\tau_t} + \boldsymbol{\phi_k} + \epsilon_{it} \quad (4.3)$$

$Novelty_{it}$ is measured by the absolute values of the Z scores for product attributes (wavelength, power, repetition rate, pulse length) and measure the divergence from the average for the focal laser's major type and year. $Soviet_Share_i$ is calculated as the share of Soviet articles out of global articles published for the focal laser's type between 1980 and 1990. $Post_1990_t$ is a dummy equal to 1 if the year is after 1990 and zero otherwise. If science reduces the technological cost to introducing novel products, I expect $\hat{\beta}_1 > 0$. τ_t refers to year fixed effects and partially mitigates concerns that year-specific shocks to laser demand caused by events such as the dotcom bubble or the global financial crisis do not systematically affect the success or failure of Soviet-type lasers. ϕ_k are fixed effects for the fourteen construction (7)-beam type (2) combinations and addresses the possibility that $\hat{\beta}_1$ merely identifies effects from comparing completely disparate laser types (e.g. a pulsed dye laser

versus a continuous wave solid state laser).[36]

Table 4.6 presents results from estimating Equation 4.3. As expected, $\hat{\beta}_1$ is positive and significant. Lasers with a standard deviation larger share of Soviet papers are 62% more novel in their wavelengths after the fall of the Berlin Wall, compared to the sample mean, even after including year and construction and beam type fixed effects.

[36]I also estimate a product-level specification for first product introductions in Column 1 of Table 4.10.

Table 4.6: Novelty of Laser Products

	DV: $\|Z\ Score\|$				
	(1) Wavelength	(2) Power/Energy	(3) Diameter	(4) Pulse Length	(5) Rep Rate
Soviet Paper Share × Year > 1990	2.428	0.772	0.757	2.174	2.759
	(0.283)	(0.369)	(0.158)	(0.771)	(0.682)
Soviet Paper Share	2.887	15.725	5.180	1.068	-1.019
	(1.077)	(1.134)	(0.295)	(0.935)	(1.065)
Average of DV	1.114	1.716	0.813	1.797	1.803
Year FEs	Yes	Yes	Yes	Yes	Yes
Construction-Beam Type FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.007	0.025	0.004	0.002
Number of Observations	65,560	66,467	41,815	17,814	21,823

Notes: Unit of analysis is the product-year. $Year$>1990 is equal to one if the panel year is larger than 1990. *Soviet Paper Share* refers to the share of Soviet publications out of global laser publications related to the product's gain medium (Weber, 2000). $|Z\ Score|$ for wavelength is calculated by dividing the difference between the laser's wavelength and the mean wavelength by the mean wavelength. The dependent variables in the other columns are calculated analogously for the respective product attributes. Standard errors are robust to arbitrary heteroscedasticity.

I verify whether there is a similar increase in novel inventions relying on Soviet science in the patent data. I regress patent novelty (measured as one minus the patent textual similarity measure outlined in Section 4.4) against patent citations to Soviet journals in Table 4.7. I control for the number of citations made to any scientific article found in MAG ($Science_Citations_i$) as citations to Soviet science may be positively correlated with the general scientific sophistication of the inventors or patent examiners searching for relevant non-patent prior art. Proxies for patent quality such as forward patent citations, triadic patent status and number of claims are also included as controls as quality may confound the novelty of the invention. I find in Column 1 of Table 4.7 that citing a Soviet laser paper is associated with around 3% higher novelty score relative to the sample mean.[37] These results suggest that newly available scientific knowledge from the erstwhile Soviet Union was exploited by Western laser firms to invent and field more novel products.

[37]I find in Appendix Table C.5 that the effect dissipates as the "degrees" of citation connection (Ahmadpoor and Jones, 2017) increase until the second degree (column 3). At the third degree, the direction of the effect flips, showing the diminishing returns to this mechanism.

Table 4.7: Novelty of Laser Patents (CPC: H01S) and Citations to Soviet Science, by Assignee Type

	DV: Patent Novelty			
	(1) All	(2) Univ/Gov	(3) Startup	(4) Compustat
Cites to Soviet Science (D=0)	0.022	0.090	0.027	0.005
	(0.010)	(0.040)	(0.024)	(0.010)
Patent Cites to Science	-0.000	-0.001	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Forward Patent Cites	-0.001	0.022	-0.002	0.001
	(0.000)	(0.008)	(0.003)	(0.001)
Triadic Patent Dummy	0.005	-0.039	0.005	0.009
	(0.002)	(0.020)	(0.008)	(0.003)
Number of Claims	0.001	0.000	0.000	0.001
	(0.000)	(0.001)	(0.000)	(0.000)
Length of First Claim	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.762	0.772	0.785	0.788
Years	34	10	28	31
R^2	0.088	0.081	0.134	0.075
Number of Observations	15,436	253	691	5,010

Notes: Unit of analysis is the patent. *Patent Novelty* is calculated as one minus the normalized textual similarity of a patent to all U.S. prior art published from 1980 (please see (Arora et al., 2018a) for details). Sample is limited to patents in CPC: H01S granted between 1980 and 2010. Sample for Column 2 consists of patents by academic assignees (including university hospitals) and government assignees (including government agencies, laboratories, and branches of the armed forces). Sample for Column 3 consists of patents assigned to small entities with under 500 employees classified by Section 41 of the U.S. patent act. Sample for Column 4 consists of patents assigned to Compustat firms matched by the DISCERN project (Arora et al., 2021b).

4.5.2 Invention (Patents)

I do not find that startups are more likely to use Soviet science than incumbents. Also, startups are not more likely to use Soviet science to develop novel inventions. In the right-hand panel of Table 4.8, I find that startups are most likely to generate

patents with optical gain media in which the Soviet Union published more in during the 1980s. However, the difference with the public sector or large firms is not statistically significant.

While the extensive margin suggests that startups patents are not more likely to be based on Soviet science, it is possible that those that do use Soviet science are more likely to produce novel patents. Therefore, I split the sample in Column 1 of the patent level regression table 4.7 by the type of assignee (university & government; small firm with under 500 employees (based on USPTO maintenance fee payments); and a listed firm in S&P Compustat). I find that the coefficient estimates for startups and Compustat firms in Columns 3 and 4 are positive but not significantly different from each other. Rather, it is the university and government subsample in Column 2 that shows the strongest correlation between exposure to Soviet science and patent novelty. Appendix Table C.4 replicates these results using the gain media collected from patent InChI codes: patents with gain media where the Soviets published more in become more novel after 1990 only for the university and government subsample (for startup and Compustat firms, the coefficient estimates are not significantly different from each other). In short, startups were not more likely to create novel patents using Soviet science. Therefore, if startup products linked to Soviet science are more novel than incumbents (as I show in Section 4.5.4), it cannot be because startups were more likely to use Soviet science. As well, these results mitigate concerns that Soviet knowledge was preferentially accessed by either startups or incumbents based on superior "fit" or prior investments in absorptive capacity.

Table 4.8: Citations to Soviet Laser Journals and Gain Media's Soviet Paper Share (H01S Patents)

	Share of Patents Citing Soviet (%)			Soviet Paper Share of Gain Medium		
	Count	Mean	Std Dev	Count	Mean	Std Dev
All	16254	1.03	10.11	11116	0.24	0.22
Government/University	1762	2.84	16.61	1277	0.25	0.23
Startup	1344	1.41	11.81	972	0.27	0.24
Compustat	4372	0.89	9.40	3027	0.26	0.23

Notes: Unit of analysis is the patent. Sample is limited to patents granted between 1990 and 2010 in the H01S CPC. The "Government/University" group refers to patents granted to assignees matched to public bodies and academia. "Startup" group consists of patents by small entities as defined by Section 41 of the U.S. Patent Act. "Compustat" patents are sourced from the DISCERN project. "Share of Patents Citing Soviet (%)" is the percentage of patents in each group that cite a paper published in Soviet quantum electronics journals listed in Section 4.4. "Soviet Paper Share of Gain Medium" refers to the share of Soviet papers for optical gain media in Weber (2000) that is matched to the InChI codes parsed in the patent text (where there are multiple media, I keep the one with the highest share).

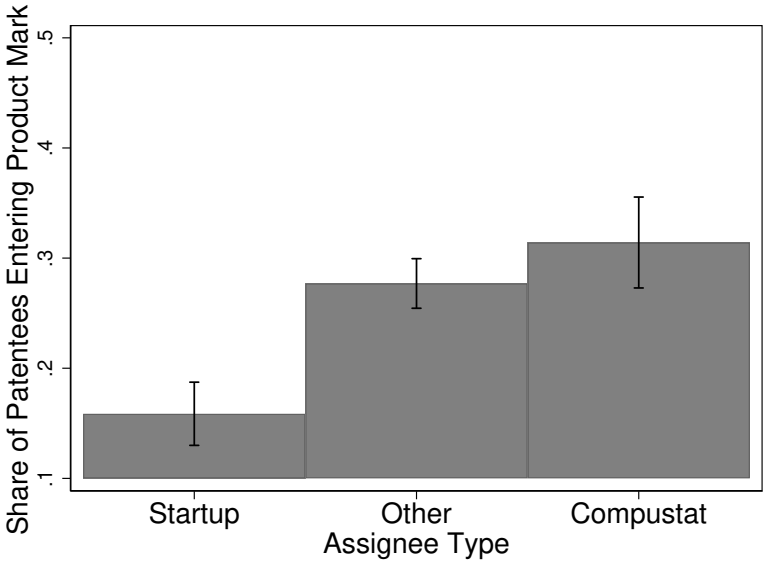


FIGURE 4.3: Entry of Patentees into Product Market[39]

4.5.3 Entry

Result 1 predicts that conditional upon patenting, startup entry rates into the downstream market should be lower than incumbents'. This is because startups face higher entry costs due to their downstream commercialization disadvantage relative to incumbents. At the assignee-year level, the probability of fielding a product in *Laser Focus* conditional upon patenting is around 21% for startups, while around 39% for Compustat firms. The 95% confidence intervals in figure 4.3 show that the difference between startup patentees and the rest of the sample is statistically significant. I replicate this result after restricting the sample to patentees in Soviet lasers[38] after 1990 (entry rate is 21% for startups and 35% for Compustat firms).

Column 1 of Table 4.9 confirms that entry rates for startups are lower for startups

[38]Those with gain media whose Soviet paper share are in the fourth quartile.

[39]This bar graph plots the average entry rates of patentees in the H01S CPC, split by the type of the assignee. Entry is defined at the assignee level if the patentee's name is found in the *Laser Focus* Buyer's Guide. The "Startup" group consists of patentees that are classified as a small entity per Section 41 of the U.S. Patent Act. "Compustat" group consists of assignees matched to the DISCERN database. "Other" group refers to all other assignees except for public entities and university assignees.

than Compustat firms, even after controlling for assignee characteristics proxying for differences in inventive quality such as average forward citations. For Columns 2 and 3, I limit the sample to assignees patenting inventions listing chemical compounds (linked by InChI codes to Weber (2000)) with higher Soviet paper shares. Startups using Soviet science are less likely to enter after 1990 (column 2), but not before 1990 (column 3). Increased technological opportunity (e.g., due to the influx of Soviet knowledge) may generate more frequent "draws" of new inventions by startups, but this may also increase the risk of "duds'; alternatively, it is also possible that the increased entry of specialized inventors may result in higher activity in markets for technology, which enables commercialization in the downstream market without entry (Arora et al., 2022a). That is, firms could choose to "cooperate" through reassigning patents or licensing rather than to "compete" in the downstream market (Gans et al., 2002). In unreported robustness checks, I find that Soviet-type patents are more often reassigned (i.e., traded) after the end of the Cold War. Conversely, around 46% of all firms found in the *Laser Focus* Buyer's guide never patent.

Table 4.9: Entry of Laser Patentees into Product Market

	DV: Entry=1		
	(1)	(2)	(3)
	All	Soviet,Post90	Soviet,Pre90
Startup	-0.108	-0.122	0.088
	(0.019)	(0.039)	(0.139)
Compustat	0.030	0.027	0.250
	(0.024)	(0.054)	(0.186)
Avg(Cites Science)	-0.001	-0.001	-0.021
	(0.000)	(0.002)	(0.025)
Avg(Forward Patent Cites)	-0.002	0.014	-0.107
	(0.001)	(0.023)	(0.083)
Avg(Number of Claims)	-0.001	-0.000	0.002
	(0.000)	(0.001)	(0.006)
Avg(Length of First Claim)	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Avg(Triadic Patent Dummy)	0.008	-0.001	-0.086
	(0.019)	(0.042)	(0.133)
Average of DV	0.256	0.237	0.114
Years	31	20	7
R^2	0.038	0.075	0.368
Number of Observations	2,631	558	35

Notes: Unit of analysis is at the assignee level. Dependent variable is equal to one if the name of the assignee appears at least once as a seller in the *Laser Focus* Buyer Guide. "Startup" is a dummy equal to one if the assignee is a small entity under Section 41 of the U.S. Patent Act. "Compustat" is a dummy equal to one if the assignee is a compustat firm matched to DISCERN. The sample excludes public entities and universities. The base group is all other private assignees. Sample for Column 2 and 3 consist of assignees whose Soviet paper share is in the 4th quartile. Column 2 consists of assignees whose first patent was granted after 1990 (and column 3, before). The Soviet paper share for an assignee-year is the average of the patent-level Soviet paper share of the InChI codes linked to Weber (2000). Fixed effects for the first year the patentee was assigned a laser patent are included. Standard errors are robust to arbitrary heteroscedasticity.

4.5.4 Innovation (Products)

Result 2 predicts that startups are more likely to commercialize novel innovations. In the sample of product-fielding firms, startups are around three times as likely

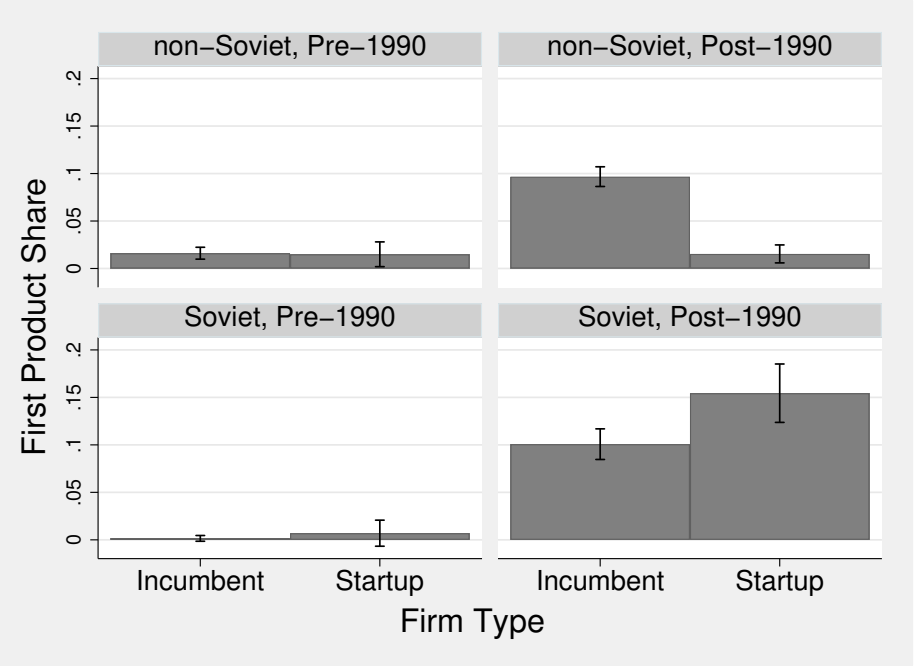


FIGURE 4.4: Product Novelty and Firm Type, Before and After Soviet Collapse[40]

to introduce first products as incumbents ($t = 8.38$). Figure 4.4 splits the product sample by Soviet paper share and before/after 1990.

The top rows show that the probability of introducing a new product both fall after 1990 for "non-Soviet" laser products. The bottom rows show that "Soviet" laser products (with gain media in the 4th quartile of Soviet paper shares) both experience a rise in first product shares. For startups introducing "Soviet" laser products, this share jumps from 10% to 15% after the end of the Cold War.

I estimate Equation 4.3 where $Novelty_i$ is equal to one if a product is the first in its type[41] and zero otherwise. I replicate the Table 4.6 results in Column 1 of

[40]This bar graph plots the share of laser products found in the *Laser Focus* Buyer's Guide that were first in their product type (defined as a construction-beam type-gain medium combination). Any subsequent product of the same type that follows the first product in later years receives a value of zero for "First Product". Sample for lower panels is limited to products whose gain media's Soviet paper shares are in the fourth quartile (the upper panel consists of all other products). The plots are split horizontally by years before and after the Cold War (sample in right panels are limited to products introduced after 1990). Startup is defined as a firm whose age (defined as panel year minus foundingyear) is five years or less.

[41]The type defined by construction-beam type-gain medium combinations. Pulsed Manganese

Table 4.10, which finds a positive and statistically significant relationship for $\hat{\beta}_1$ in Equation 4.3. To test the prediction that this science-induced commercialization of novel products is driven by startup entry, I split the sample in Columns 2 and 3 by whether the producer of the laser is a startup or an incumbent[42] and find that $\hat{\beta}_1$ is 4.2 times larger for startups. My unit of analysis is at the product level (not product-year level) such that a firm's product only appears once on the year when it is introduced. Given that novelty is measured as the share of new products out of total products, this prevents successful legacy products (that are continued for multiple years) from inflating the number of non-first products. Moreover, I include controls for firm size by the number of employees and find that the results hold for a subset of firms with less than 50 employees in Column 4. Conditional upon introducing a product to the market, startups are more likely to commercialize technically novel products than incumbents, which is consistent with Result 2.

gas lasers are the first such products in my sample (introduced in 1982); continuous wave Thulium fiber lasers are the last of such products (introduced in 2010).

[42]A startup is defined as a firm that is less than five years past its founding year at the time of listing the product on *Laser Focus*

Table 4.10: Startup Entry into Novel Products

	DV: First Product = 1			
	(1) All	(2) Incumbent	(3) Startup	(4) Startup(¡50)
Year > 1990 × Soviet Paper Share	0.339	0.285	1.198	1.263
	(0.046)	(0.049)	(0.116)	(0.123)
Soviet Paper Share	-0.006	0.022	-0.085	-0.060
	(0.019)	(0.022)	(0.041)	(0.048)
Employees	0.000	0.000	0.001	0.002
	(0.000)	(0.000)	(0.000)	(0.001)
Average of DV	0.077	0.079	0.070	0.062
Year FEs	Yes	Yes	Yes	Yes
R^2	0.330	0.402	0.350	0.354
Number of Observations	4,836	3,739	1,097	996

Notes: Unit of analysis is the product. *Soviet Paper Share* refers to the share of Soviet publications out of global laser publications related to the product's gain medium (Weber, 2000). *Employees* refers to the number of employees of the focal firm at the year the product is introduced. Standard errors are robust to arbitrary heteroscedasticity.

I test whether startup commercialization of novel products is solving fundamentally new problems or improving approaches to existing problems. Figure 4.5 presents evidence that "Soviet-type" products were not only technologically novel but also addressed new downstream markets. Limiting the sample to Soviet-type products introduced after 1990, I find that startup products were sold into around 1.6 new applications (new application is defined as a market that sees more than a 200% sales jump in the *Laser Focus* annual market reviews) while incumbents entered into around 40% less applications.

Table 4.11 splits the sample in Column 1 of Table 4.10 by sales jumps in new applications and downstream market concentration measures (HHI). If startup entry indeed targeted fundamentally novel problems, I expect entry to occur in new applications and in downstream markets with higher diversity (lower Herfindahl indices). Column 1 first replicates the split-sample results in Table 4.10 by interacting the difference-in-differences coefficient with the dummy for whether the laser's pro-

ducer was a startup. As expected, the triple difference coefficient is positive and statistically significant.

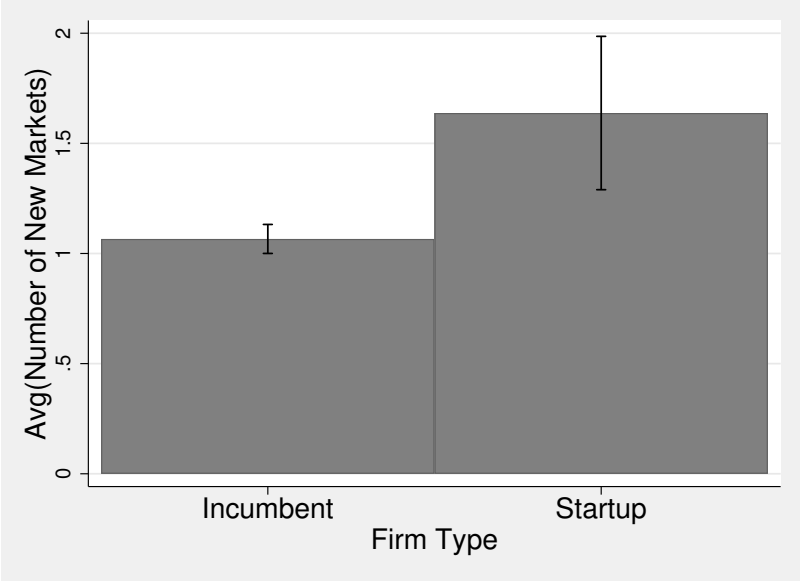


FIGURE 4.5: New Downstream Applications by Firm Type (Post-1990, Soviet-type Products)[43]

Columns 2 and 3 split the sample by whether the focal laser has seen more than a doubling in sales in at least one of the thirteen application areas outlined in Section 4.4.[44] The triple interaction coefficient in Column 2 is around 25 times larger in magnitude compared to that in Column 3, which is consistent with lasers with startups innovating in new applications. Columns 4 and 5 split the sample by whether the laser has wide or narrow applications in downstream markets.[45] The triple interaction coefficient Column 4 is around four times in magnitude compared to that in Column 5. This implies that lasers whose applications are concentrated

[43]This bar graph plots the average number of new applications for lasers with Soviet paper shares in the fourth quartile (Weber, 2000) introduced after 1990. The firm type definitions are identical to those in figure 4.4

[44]Around 70% of our product-years do not have a new application under this definition.

[45]Wideness/narrowness of applications is measured by average HHI measure of sales shares across thirteen applications for a laser between 1986 and 2005. The HHI measures are calculated for each year and then averaged over the 20 year period.

in fewer sectors tend to see relatively less startup introductions of novel products compared to lasers spread across diverse applications, after the influx of Soviet laser science. These results suggest that startups commercializing Soviet laser science did so in fundamentally novel areas rather than existing applications.

Table 4.11: Startup Entry into Novel Products, by Downstream Conditions

DV: First Product = 1	Baseline	Sales Jump		Sales HHI	
	(1)	(2)	(3)	(4)	(5)
	All	>200%	<200%	Low	High
Year > 1990 × Soviet Paper Share × Startup	1.086	2.119	0.086	1.794	0.459
	(0.114)	(0.142)	(0.176)	(0.168)	(0.164)
Soviet Paper Share × Startup	-0.089	-0.000	-0.009	-0.029	-0.121
	(0.038)	(0.001)	(0.006)	(0.037)	(0.075)
Year > 1990 × Startup	-0.365	-0.781	-0.054	-0.635	-0.165
	(0.034)	(0.046)	(0.050)	(0.050)	(0.053)
Year > 1990 × Soviet Paper Share	0.212	-0.167	-0.210	-0.016	-0.301
	(0.050)	(0.073)	(0.066)	(0.060)	(0.175)
Soviet Paper Share	0.013	-0.001	0.007	0.020	0.054
	(0.022)	(0.001)	(0.004)	(0.020)	(0.045)
Startup	0.010	0.001	0.000	-0.007	0.022
	(0.012)	(0.001)	(0.001)	(0.005)	(0.024)
Employees	0.000	0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.082	0.129	0.034	0.111	0.049
Year FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.362	0.655	0.253	0.576	0.278
Number of Observations	4,836	1,985	1,396	2,549	1,563

Notes: Unit of analysis is the product. Dependent variable is a dummy equal to one if product is the first of its type. *Startup* is a dummy equal to one if the producing firm's age is five years or less. Other variable definitions are identical to those in Table 4.10. Column 2 consists of lasers that have at least one out of thirteen application areas that exhibits a sales jump in excess of 200% in the year following the product introduction year. Column 3 consists of those observations who do not exhibit such a sales jump in a new application. *HHI* is calculated at the laser type level as the Herfindahl-Hirschman Index of the sales shares across thirteen application areas found in the *Laser Focus World* Annual Market Reviews. Standard errors are robust to arbitrary heteroscedasticity.

4.5.5 Exit

Finally, exit (failure) rates are lower for startups commercializing Soviet lasers compared to incumbents. I calculate 3,5 and 10 year survival rates after introducing a

product for firms in the Buyer's Guide. I find that 59% of startups that commercialize Soviet-type products after 1990 survive after 3 years (i.e., still field products in the same category) while 39% of incumbents survive in the same duration (the difference is statistically significant with t-statistic of 4.4).[46] Similar patterns obtain for 5 and 10 year survival rates.

Table 4.12 shows that startups commercializing Soviet science are less likely to exit than incumbents. Column 1 compares the 3-year survival rates of producers of Soviet-type lasers after 1990 and finds that startups are around 36% less likely to exit (i.e., cease selling the product in the Buyer's Guide) within 3 years of product introduction, relative to the sample mean. Startups also have higher 5-year (54%) and 10-year (137%) survival rates relative to the sample mean. Interestingly, these results do not obtain for Soviet laser products introduced before 1990 (columns 2, 4 and 6). The finding that startups are less likely to enter conditional upon patenting and exit conditional upon entering after the arrival of Soviet science is consistent with a higher entry threshold (e.g., due to a lack of downstream commercialization capabilities) for startups than incumbents: this accords with the lower share of startup patentees entering the Buyer's Guides compared to incumbents; as well, startup entrants that do commercialize in the product market are likely to have done so with high-payoff (novel) products and markets.

[46]These results also hold for all lasers, including pre-1990 and non-Soviet-type lasers.

Table 4.12: Survival Rates for Producers of "Soviet" Lasers

Dependent Variable	3yr Survival=1		5yr Survival=1		10yr Survival=1	
	(1)	(2)	(3)	(4)	(5)	(6)
	Post90	Pre90	Post90	Pre90	Post90	Pre90
Startup	0.181	0.162	0.193	-0.032	0.271	0.158
	(0.079)	(0.127)	(0.088)	(0.132)	(0.086)	(0.129)
Employees	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.497	0.638	0.359	0.478	0.198	0.304
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.169	0.073	0.154	0.081	0.215	0.114
Number of Observations	175	69	142	69	96	69

Notes: Unit of analysis is the producer-product type (construction-beam type-gain medium combination). Sample is limited to producers that sell a product in *Laser Focus* Buyer's Guides whose gain media's Soviet paper share is in the 4th quartile. Dependent variable for Columns 1 and 2 is equal to one if the firm and product type continues to appear three years after the product introduction year. Columns 1 and 2 are split by whether the introduction year of the product is after or before 1990. Columns 3,4 and 5,6 are defined analogously for 5 year and 10 year survival rates. *Employees* refer to the number of employees at the year of product introduction for the focal firm. Fixed effects are included for the introduction year for the product. Standard errors are robust to arbitrary heteroscedasticity.

4.6 Discussion and Conclusion

Whether and why startups are more innovative than incumbent firms has received increased interest in recent years. With the decline of large corporate labs that were the typical agents translating science to innovation (Arora et al., 2018b, 2020), the literature has examined alternative organizational forms such as university spinoffs (Perkmann et al., 2021; Rothaermel et al., 2007), venture capital-funded startups (Lerner and Nanda, 2020) and the mission-oriented "ARPA" model of innovation (Azoulay et al., 2019; Howell, 2017). As well, the recent productivity decline documented by Gordon (2012) and Bloom et al. (2017) and the dominance of select large firms (Autor et al., 2020; Gutiérrez and Philippon, 2020) in the economy have renewed interest in the "neo-Schumpeterian" questions linking firm size, youth and innovation productivity (Cohen, 2010). Conventional wisdom from archetypal Silicon Valley "garage ventures" and large-scale R&D productivity data (measured typically by patents per R&D dollars) show startups to be more innovative than incumbents (Griliches, 1990). However, empirical tests of theories explaining startup propensity for novel innovations rely on large scale cross-industry data that typically cannot control for technological opportunity across firms (Cohen and Klepper, 1996). For instance, startups formed by university scientists may possess superior inventive capabilities that are inaccessible to incumbents; conversely, incumbents may have capabilities built up from prior operations that can be re-purposed (Helfat, 1997). This has hampered progress to answering why startups sometimes out-innovate incumbents with new scientific knowledge.

This paper uses a shock to scientific knowledge in the laser industry caused by the end of the Cold War to answer why startups and incumbents commercialize new science differently. I find that startups in the West were more likely than incumbents to commercialize new innovations in new markets using Soviet scientific knowledge

made available after the Berlin Wall fell. However, pre-entry patenting activity shows startups and incumbents to use Soviet science at similar rates. Therefore, the post-entry product market patterns cannot be explained by startups' superior capabilities in absorbing new science. Rather, startups seem more willing than incumbents to undertake novel innovations that are more uncertain but promise higher payoffs. Consistent with this theory, startups are less likely to enter conditional upon patenting, and less likely to exit upon entering the product market. Scholars have conceptualized startups as "economic experiments" (Nelson, 1981; Rosenberg, 1994; Kerr et al., 2014; Klepper, 2015), arguing that an innovation ecosystem consisting of smaller firms can exhibit higher technological diversity and faster technological change (Cohen and Klepper, 1992). This paper lends empirical support to these prior works and tests the entry threshold mechanism to explain why startups "experiment" even if they have similar pre-entry capabilities.

Few studies have compared startups' and incumbent firms' propensities to engage in novel product introductions, while those that do fail to account for potential underlying differences in inventive capabilities (Prusa and Schmitz Jr, 1991). More recent work has identified contingencies under which firms may attempt more novel products (Eggers and Park, 2018), but focuses primarily on pre-entry experiences that build up upstream capabilities (Helfat and Lieberman, 2002). The literature has also studied implications of specialized downstream complementary assets such as sales, but mainly concluded the first order effect on whether firms will enter, without considering what kinds of innovations such entrants will bring (Mitchell, 1989; Tripsas, 1997). Similarly, the technology commercialization framework by Teece (1986) and Gans et al. (2002) and the division of innovative labor literature (Arora et al., 2004; Gans and Stern, 2010) also incorporate downstream appropriability conditions, but the main consequence from higher downstream entry costs for startups (in the form of weaker IP regimes or incumbent complementary asset investments) is on the deci-

sion to "cooperate" versus "compete" with incumbents, rather than on the types of innovations to "compete" with. My results that higher entry thresholds can increase the novelty of innovations brought by startups is therefore a novel contribution to the literature.

The immediate implications of this study bear on innovation policies on entrepreneurship and "big-tech" regulation. First, the entrepreneurship literature has explored various ways to encourage new firm entry by reducing institutional barriers (Klapper et al., 2006; Chang and Wu, 2014; Skiti, 2020) and improving access to financing (Hall and Lerner, 2010; Kerr and Nanda, 2015). The assumption implicit in these findings is that the entry of new entrepreneurs introduces superior products and thereby disciplines complacent incumbents (Gilbert and Newbery, 1982; Gilbert, 2006). Indeed, inventors founding new firms seem to produce riskier and more novel inventions than their colleagues at their previous jobs (Ewens and Fons-Rosen, 2013). Since lowering entry barriers encourages entry of startups that innovate more novel products, it may follow that eliminating entry costs would hasten the translation of scientific discoveries into innovation. My findings caution against such optimism, as entries subsequent to the "superstar" startup should exhibit diminishing novelty. At the extreme, with no entry costs, startup and incumbent innovations would only reflect pre-entry capability differences. Conversely, technological changes such as the rise of big data analytics for software firms that disproportionately favor incumbents with existing resources imply that the average novelty of startup entrants must increase even further. Indeed, comparing across industries, startups that pioneer revolutionary products in commercial space exploration or electrical vehicles do so in sectors with high entry costs. Meanwhile, lower entry costs resulting from the spread of personal computing and new web browsers coincided with the proliferation of "me-too" web services in the years leading into the dot-com crash.

Second, regulations targeted specifically against "big-tech" entry into new mar-

kets may ironically spur the novelty of their innovations. The Department of Justice ruled against DuPont's operations in smokeless powders and barred AT&T from entering into computers following its transistor invention. Yet, the former responded by pioneering the new field of polymer chemistry to bring nylons to market, while the latter manufactured America's first commercial satellite (the TELSTAR), invented the C programming language and pioneered modern fiber optic communications.

There are other arguments for encouraging firm entry that are beyond the scope of this paper. Since my theoretical framework does not encompass product market interactions between producers, it overlooks the role of competition. The entry of more firms may incentivize incumbents to lower prices to preempt further entry, allowing for greater value capture by consumers (Klepper, 1996). To the extent that some consumers can substitute products with differing novelty, the less novel, marginal entrant can still trigger this process. Hence, future work may distinguish between product and process innovations to paint a fuller picture of how science impacts an innovation ecosystem.

5

Conclusions

> "Science, by itself, provides no panacea for individual, social, and economic ills. It can be effective in the national welfare only as a member of a team, whether the conditions be peace or war." (Bush, 1945)

The introduction has motivated this dissertation by juxtaposing the continued rise in American scientific output with lagging productivity figures in recent years. The twin trends underline that while the flow of scientific knowledge is a necessary condition to economic growth, the delivery of its fruits in the form of new innovations requires a careful understanding of firms that remain the key protagonists in the application of scientific knowledge to innovations. The three studies of this dissertation are attempts to unpack the mechanisms that drive corporate participation in the creation and exploitation of scientific knowledge for innovation.

The first study shows that science affects innovation indirectly by promoting markets for technology. By facilitating an efficient allocation of existing inventions to the best user and the entry of new innovators capitalizing on this prospect, science may hasten its own commercialization.

The second study shows how the unavailability of public science may incentivize firms at the technological frontier to invest more in their own research. If this is true, then more public investments in university science may ironically dampen incentives for scientific research at firms on the technological frontier, even as innovators in general benefit from the greater availability of scientific knowledge.

The third study shows that a greater abundance of scientific knowledge need not translate to novel innovation, and that the direction of technical change is contingent upon the types of entry costs faced by potential innovators. The results gleaned from laser firms that responded to the abrupt availability of Soviet laser science suggests that startups which face higher entry costs decide to "swing for the fences" with novel, first-in-kind lasers.

The dissertation therefore examines the relationship between science and three distinct elements: i) markets for technology; ii) corporate research; and iii) innovation novelty. Future research may explore the interactions between these three elements. For instance, what is the nature of the connection between markets for technology and innovation novelty? Chapter 2 argues that active MFTs can encourage the entry of new science-based firms. Will such firms enter with innovations that are novel or incremental? Even as the literature has studied the effects of general purpose technologies on a division of innovative labor between firms (Bresnahan and Gambardella, 1998), the obverse question of how a division of innovative labor affects the direction of technical change is a relatively understudied question. Indeed, the startups that patent but do not enter the product market in Chapter 4 may well be specializing in more incremental patents that are intended to be sold or licensed to incumbent firms. Such a division of innovative labor would help explain the lower entry rates of startups upon patenting and lower novelty of incumbents in the product market. Given that the division surveyed in Chapter 1 typically envisions the startup as mining the novel inventions that are then commercialized by the incum-

bent firm, the possibility that markets for technology are principally responsible for the sourcing of incremental inventions may be a surprising result to examine further.

Appendix A

Appendix to Chapter 2

A.1 Data

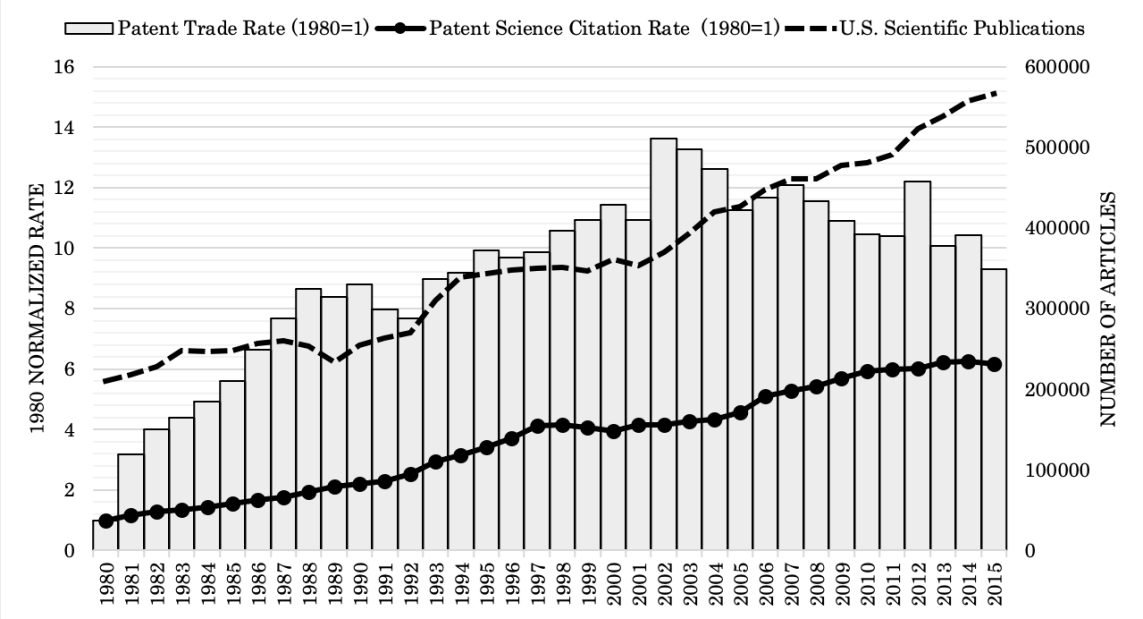


FIGURE A.1: Markets for Technology and Production and Use of Science[1]

A.1.1 Sample Construction

Our patent data is from the 2016 publication of PatStat and encompasses around 5.2 million utility patents granted by the USPTO from 1980 to 2016. We collect information on patent reassignment (transaction date, identity of buyers and sellers) by linking them to the USPTO Patent Assignment Database (PAD) (Graham et al., 2018), which records details on the transfer of ownership between patent assignees. To account for sample truncation, we limit our sample to patents granted on or before

[1]This graph plots time trends of patent trade, patent citations to science, and U.S. scientific publication output over our sample period. *Patent Trade Rate* for a given year is defined as the ratio between the number of patents reassigned over the number of patents in force in that year. *Patent Science Citation Rate* for a given year is defined as the number of citations to scientific articles in Microsoft Academic Graph (MAG) divided by the number of patents published in that year. Both rates are normalized by their levels in 1980 in this graph. *U.S. Scientific Publications* refer to the total number of scientific publications with a U.S. author in a given year from Clarivate Web of Science's Science Citation Index-Expanded (SCI-EXPANDED) and Conference Proceedings Citation Index-Science (CPCSI-S).

2011 (for which we observe reassignments until 2015).[2] We construct measures of invention novelty based on textual similarity and on technological combinations, following Fleming (2001).

Patent Citations to Science

We employ a publicly available dataset from Marx and Fuegi (2020), which matches NPL citations to scientific articles available in Microsoft Academic Graph (MAG). The dataset assigns confidence scores for matches between a patent's NPL citation and a MAG article (1 being the lowest and 10 being the highest). We take the "PCS (Patent Citations to Science)" file and first exclude matches with under a 90% confidence score. We further exclude cited articles in the social sciences or humanities, leaving us with OECD subject fields in "Natural Sciences", "Engineering and Technology", "Medical and Health Sciences", "Agricultural Sciences".

Textual similarity to science — We take the pairwise textual similarity measure from Arora et al. (2018a), which calculates cosine similarities on the text of U.S. patents and scientific articles from the Science Citation Index - Expanded collection of Clarivate Web of Science. For each paper published between 1990 and 2015, we sort the patents in descending order of their similarity scores to the focal paper. We then rank the top 100 patents in terms of similarity scores to each publication ("similar patents"). The *Textual Similarity to Science* variable is equal to the natural log of one plus the number of publications for which a patent is classified as a "similar patent". In unreported robustness checks, we calculate two additional metrics. First, we normalize the the similarity scores by the maximum similarity score pair for a publication, and take the average of this score for each patent. Second, we take the average of the similarity rank each patent receives with respect to each publication.

[2]About 58% of patents that are reassigned are done so within five years of being granted.

The direction and statistical significance of the column 6 result in table 2.3 are not sensitive to these alternative calculations of patent textual similarity to science.

Journal Impact Factor — Journal impact factor for a journal in year t is calculated as the number of forward citations in years $t-1$ and $t-2$ received by the number of papers published in years $t-1$ and $t-2$ by the focal journal. The average journal impact factor for patent i averages this value for all articles cited by a patent.[3]

Citation Lag — We measure how recent the science being cited is in relation to a patent by measuring the average year difference between the grant year of the patent and the publication year of the paper. For patent i citing $j \in J_i$ articles,

$$Avg(Lag\ to\ Cited\ Science_i) = \frac{\sum_{j \in J_i} Grant\ Year_i - Publication\ Year_j}{|J_i|}$$

The lower this value, the more recent ("younger") the science being used in relation to the patent.

Number of Fields — we calculate how specialized a patent's scientific citations are by counting the number of unique WoS Fields of cited scientific papers per patent and dividing it by the number of cited papers (the measure ranges between 0 and 1). For patent i citing J articles published in K fields,

$$Normalized\ Field\ Counts_i = \frac{|K_i|}{|J_i|}$$

The lower (higher) this value, the more specialized (interdisciplinary) the patent is in terms of its scientific citations.

[3]We source journal impact factors from Marx and Fuegi (2020)

Scientific Combination Familiarity — we calculate the novelty of the combination of Web of Science scientific fields cited by counting how many times the same scientific combinations have been cited by the paper cited by a focal paper since 1790.

$$Combination\ Familiarity\ for\ WOS\ Fields\ (Decayed)_i = \sum_{patents\ k\ before\ patent\ i} 1\{k\ cites\ identical\ combination\ of\ WOS\ Fields\ as\ i\} \times exp(\frac{grant\ date\ of\ patent\ k\ -\ grant\ date\ of\ patent\ i}{time\ constant\ of\ knowledge\ loss})$$

Where the *time constant of knowledge loss* is set to 5 years such that a previous Web of Science combination from five years ago is weighted by $exp(-\frac{1}{5}) = 37\%$. This is an analog of the Technological Combination Familiarity measure by Fleming (2001), which is calculated for patent classes. There are 175 Web of Science (WoS) Fields assigned to 1,740,815 articles cited by patents in our sample. Intuitively, the more often the same combination appears (the higher the *Familiarity* score), the less novel are the patent's scientific combinations.

Identifying market transactions for patents

We download the 2016 version of the USPTO Patent Assignment Dataset and identify patent reassignments that may classify MFT transactions. Our framework follows methods pioneered by Serrano (2010) and refined by Ma et al. (2017) and Figueroa and Serrano (2019).

We define MFT transactions as transfers of technology between two independent entities. This excludes ownership transfers within firms and purchases of capabilities rather than technology (e.g. M&As that transfer lab personnel and capital equipment along with patents). The USPTO records each received patent transfer in a "Reel Frame" (RF) ID, and has classified the conveyance types of these transfers into

assignment of assignor's interest, name changes, government interest agreements, security agreements, and release by secured parties. We exclude all other conveyance types than assignments of assignors' interest. The USPTO also identifies employer assignment as the first recorded transaction for a patent where the patent is transferred alone with an execution date prior to the patent application disposal date (Graham et al., 2018, p.27). These RF IDs are also removed.

We add several additional checks. First, we exclude assignments whose date is before the grant date of a patent. While it is possible that a transaction has occurred before the patent was granted, it is also possible that the patent's initial assignment was mistaken with a reassignment to a buyer. Without a way to positively identify pre-grant patent application purchases, we decide it is safer to exclude these cases to reduce false positives.

Second, we exclude cases where the assignee ("buyer") names in the PAD records are similar to assignee names in the USPTO PATSVIEW. The assignee names in PATSVIEW record the initial assignee name(s) on the granted patent document. Therefore, if the assignee name in the PAD records are similar to the original owner's (assignee on patent document), we can rule out an MFT transfer between two independent entities.

Third, we exclude cases where the assignor ("seller") of an assignment is similar to the inventor of the patent from USPTO PATSVIEW. These cases are likely to be corporate employees transferring their patent rights to their firms per terms in their employment contract (it has been common practice among large corporations such as Du Pont, IBM, and Google to automatically transfer patent rights from employees to employers by such contracts).

Fourth, we download all completed acquisitions recorded in SDC Platinum between 1980 and 2015 and match the "Target Name" and "Acquiror Name" in SDC to patent assignor and assignee names in PAD. If the buyer-seller pair of companies

in SDC correspond to the buyer-seller pairs in PAD, we exclude them.

Fifth, we also measure the string distance between assignor-assignee pairs so that intra-corporate reassignments (from, say, a company's headquarters to its licensing subsidiary) are dropped. For the second to fifth steps, we judge that names are similar based on Jaro-Winkler, Jaccard, and a normalized Levehnstein edit distance (python package available from https://github.com/seatgeek/fuzzywuzzy) after standardizing common suffixes such as "CORP", "LTD" and prefixes such as "LEGAL REPRESENTATIVE". Specifically, we take one minus the maximum value of the distance measures (which range between zero and one) and classify those pairs with larger than an appropriate threshold as similar to each other. We conduct extensive human checks around these thresholds to reduce classification error.

Sixth, we exclude RF IDs with more than 25 patents being transferred, because these are likely to be part of M&A deals between large firms.

Technological sector classifications, By 2 digit IPCs

Table A.1: Technological Sector Classifications, by 2 Digit IPCs)

	Technological Sector				
2digit IPCs	Life Sciences	Chemicals	ICT	others	Total
A0	0	21,088	0	33,988	55,076
A2	0	19,613	0	8,672	28,285
A4	0	0	0	80,474	80,474
A6	260,352	13,368	0	55,149	328,869
B0	0	97,147	0	15,620	112,767
B2	0	0	0	156,854	156,854
B3	0	19,394	0	5,506	24,900
B4	0	0	0	72,225	72,225
B6	0	0	0	260,836	260,836
B8	0	0	0	1,486	1,486
C0	15,111	229,919	0	59,285	304,315
C1	53,768	29,920	0	43	83,731
C2	0	28,317	0	16,968	45,285
C3	0	5,038	0	0	5,038
C4	0	625	0	0	625
D0	0	1,971	0	27,061	29,032
D2	0	0	0	9,512	9,512
E0	0	0	0	68,229	68,229
E2	0	0	0	29,611	29,611
F0	0	0	0	104,165	104,165
F1	0	0	0	110,113	110,113
F2	0	4,985	0	66,463	71,448
F4	0	0	0	15,259	15,259
G0	26,549	0	489,134	273,734	789,417
G1	0	0	51,524	8,890	60,414
G2	0	0	0	9,437	9,437
H0	1,511	2,592	480,584	214,912	699,599
Total	357,291	473,977	1,021,242	1,704,492	3,557,002

Notes: This table tabulates the four technological sector classifications (Life Sciences, Chemicals, ICT, and other) that we use in our main sample of 3.5 million USPTO patents published between 1980 and 2011.

A.1.2 Instrumental Variable Construction

Data collection

We collect federal procurement contracts for research and development services by manually downloading government contracts by year and agency from https://www.

usaspending.gov and https://beta.SAM.gov. The former is mandated by the Federal Funding Accountability and Transparency Act of 2006 and is maintained by the Office of Management and Budget and includes all procurement activities of the U.S. federal government since 2000. The latter is run by the General Services Administration and contains procurement data as early as the 1970s. We keep data from 1980 onwards for our analysis.

Table A.2: Government Funding of Procurement Contracts for R&D Services

Rank	PSC	Contract Value 1986-88	Contract Value 1990-1992	Description
1	AN41	$ 249,686	$ 99,000,000	R&D- Medical: Health Services (Basic Research)
2	AH96	$ 143,921	$ 25,000,000	R&D- Environmental Protection: Other (Management/Support)
3	AE21	$ 253,278	$ 27,800,000	R&D- Economic Growth: Product/Service Improvement (Basic Research)
4	AN12	$ 4,530,456	$ 315,000,000	R&D- Medical: Biomedical (Applied Research/Exploratory Development)
5	AN11	$ 26,600,000	$ 1,420,000,000	R&D- Medical: Biomedical (Basic Research)
6	AN46	$ 47,000,000	$ 1,970,000,000	R&D- Medical: Health Services (Management/Support)
7	AG94	$ 851,749	$ 32,600,000	R&D- Energy: Other (Engineering Development)
8	AN15	$ 313,241	$ 9,018,358	R&D- Medical: Biomedical (Operational Systems Development)
9	AE33	$ 2,653,725	$ 63,200,000	R&D- Economic Growth: Manufacturing Technology (Advanced Development)
10	AE35	$ 17,300,000	$ 386,000,000	R&D- Economic Growth: Manufacturing Technology (Operational Systems Development)
...	...	...	...	...
212	AR12	$ 11,200,000	$ 553,805	R&D- Space: Aeronautics/Space Technology (Applied Research/Exploratory Development)
213	AS21	$ 1,865,031	$ 65,330	R&D- Modal Transportation: Surface Motor Vehicles (Basic Research)
214	AH32	$ 2,703,422	$ 69,106	R&D- Environmental Protection: Water Pollution (Applied Research/Exploratory Development)
215	AR22	$ 54,400,000	$ 1,283,886	R&D- Space: Science/Applications (Applied Research/Exploratory Development)
216	AG73	$ 3,824,674	$ 54,971	R&D- Energy: Solar/Photovoltaic (Advanced Development)
217	AG55	$ 17,900,000	$ 144,088	R&D- Energy: Nuclear (Operational Systems Development)
218	AN40	$ 10,600,000	$ 62,824	R&D- Health Services
219	AD54	$ 50,600,000	$ 189,912	R&D- Defense Other: Fuels/Lubricants (Engineering Development)
220	AR94	$ 881,000,000	$ 775,323	R&D- Space: Other (Engineering Development)
221	AZ10	$ 368,000,000	$ 39,265	R&D- Other

Notes: The observations are sorted in descending order by the logged difference between the pre (1986-88) and post (1990-92) period.

Our procurement data covers all contracts signed by the Department of Defense, Energy, Health and Human Services, and Veterans Affairs between 1980 and 2020. As of FY2019, these agencies accounted for more than 72% of all procurement contracts and constitute four of the five largest spenders on "contractual services and supplies" (the omitted agency is the Office of Personnel Management, which primarily deals with health benefits and life insurance funds and unlikely to contract out R&D services). We collect the signing date, action obligation[4] in current dollars, vendor names, contracting agency and the relevant 4-digit Product and Service Codes (PSC) for these four agencies between 1980 and 2019. We match the vendor names to the firm names found in the DISCERN database of American public, R&D performing firms between 1980 and 2015, using a combination of automated string distance metrics and manual cleaning. We limit procurement contracts to services related to R&D only (1st digit PSC corresponding to "A").

Crosswalk definition

We crosswalk the values of the contracts from PSC codes to patent classes (IPCs) and publication fields (WoS fields) using the following methods:

Crosswalk from PSC to IPC — We calculate the level of federal scientific spending relevant to an IPC by multiplying a dyadic weight based on patent class-level patenting distributions of vendor firms winning contracts in a 4-digit PSC. For firm i patenting in IPC k and contracting in product code j, the 4-digit PSC-to-4-digit

[4]Action obligations are "intentions" backed by a contractual agreement. The government does not release actual dollar amounts spent on contracts. After initial contract signing, the actual expenditure can increase, decrease, or stay the same. Agencies enter a "contract action" into the database whenever they know that what was initially obligated has changed. These corrections may sometimes (5%) lead to negative obligations.

IPC weight is defined as:

$$weight_{jk} := \sum_i Contract_Value_{ij} \times \frac{patents_{ik}}{patents_i} \tag{A.1}$$

for all DISCERN patents and contracts published and signed between 1980 and 1992. The post Cold War funding shock relevant for each 4-digit IPC (used in the first stage regression of table 2.10 column 5 and appendix table A.3) is calculated as the logged difference between the average contract value for a "pre" period from 1986 to 1988 and a "post" period from 1990 to 1992.

$$\Delta ln(Gov.\ R\&D\ Contracting_k) := ln\left(\frac{\sum_j weight_{jk} \times \frac{\sum_{t=1990}^{t=1992} Contract_Value_{jt}}{3}}{\sum_j weight_{jk} \times \frac{\sum_{t=1986}^{t=1988} Contract_Value_{jt}}{3}}\right) \tag{A.2}$$

Crosswalk from PSC to WoS Field — We calculate the level of federal scientific funding relevant to a Web of Science field by multiplying a dyadic weight based on scientific field-level publication distributions of contracting vendor firms. For firm i publishing in WoS field l and contracting in product code j, the 4-digit PSC-to-WoS field weight is defined as:

$$weight_{jl} := \sum_i Contract_Value_{ij} \times \frac{papers_{il}}{papers_i} \tag{A.3}$$

for all DISCERN papers and contracts authored and signed between 1980 and 1992.

Number of Predicted Papers — Using the above crosswalk, we run an OLS regression to predict the number of papers as a function of the funding shocks around the end of the Cold War:

$$\begin{aligned} Number\ of\ Papers_{lt} = \beta_0 + \beta_1 \Delta ln(Government\ R\&D\ Contracting)_l \\ + \beta_2 ln(Government\ R\&D\ Contracting\ (Pre))_l + \boldsymbol{\xi_t} + \nu_{lt} \end{aligned} \tag{A.4}$$

where $\boldsymbol{\xi_t}$ are paper publication year fixed effects. The WoS field level funding shock in equation A.4 is defined as:

$$\Delta ln(Gov.\ R\&D\ Contracting_l) := ln\left(\frac{\sum_j weight_{jl} \times \frac{\sum_{t=1990}^{t=1992} Contract_Value_{jt}}{3}}{\sum_j weight_{jl} \times \frac{\sum_{t=1986}^{t=1988} Contract_Value_{jt}}{3}}\right) \quad (A.5)$$

Crosswalk from WoS Field to IPC — To construct the instrument for columns 3 and 4 of table 2.10, we crosswalk the number of predicted papers ($\widehat{Number\ of\ Papers_{lt}}$) at the web of science-year level to the patent class-year level by multiplying a dyadic weight based on the NPL citations from patents the scientific literature. For patents in 4-digit IPC k citing papers in WoS field l, the number of scientific papers relevant for each IPC is defined as:

$$Number\ of\ Papers_{kt} = \sum_l weight_{kl} \times Number\ of\ Papers_{lt} \quad (A.6)$$

where the weight is defined as:

$$weight_{kl} = \frac{NPL\ Citations_{kl}}{Total\ NPL\ Citations\ Received_k} \quad (A.7)$$

for all patents granted between 1980 and 1992.

Comparison to other sources of R&D expenditure data

We use the National Science Foundation's "National Patterns of R&D" data series to compare the relative magnitude of federal R&D funding to other sources and verify that the procurement data we use for the construction of the instrument. Figure A.2 shows that the federal government was responsible for funding around 43% of all R&D expenses between 1986 and 1992. Moreover, the R&D procurement data we are using for the instrumental variable covers around 76% of total federal R&D spending performed by industry.

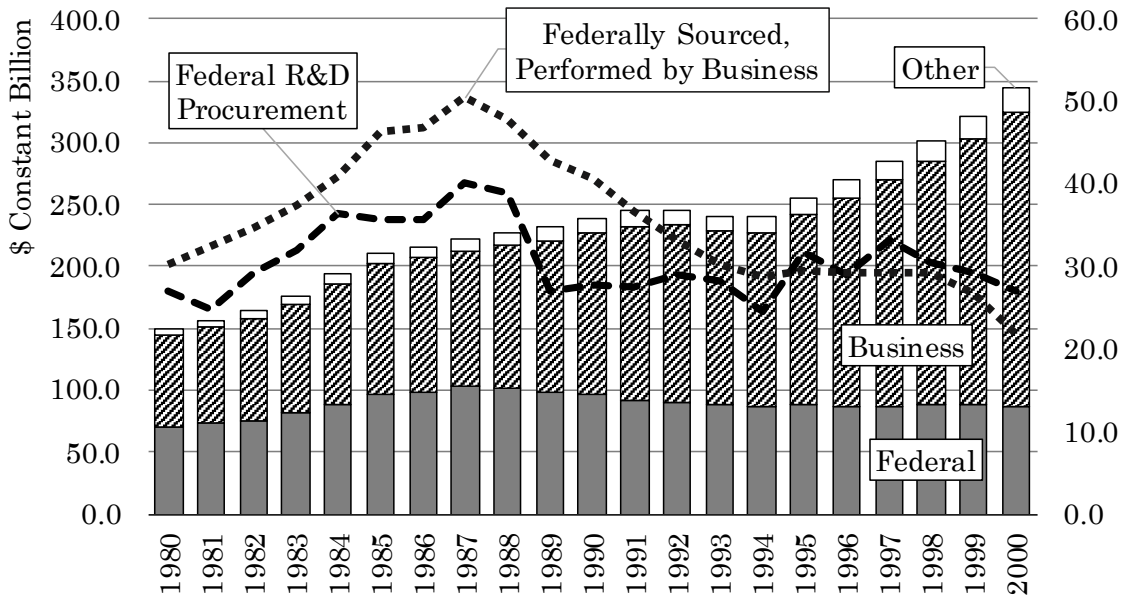


FIGURE A.2: U.S. R&D Funding, by Source[5]

There are three likely sources of this discrepancy. First, we collect data for the four largest spenders on R&D, while the NSF data covers all federal agencies. Second, reporting behaviors between the procurement data and the survey data tend to differ (Pece, 2016). That is, even for the same agency and subcategory of spending, there are significant gaps in the dollar values reported between the procurement data and the NSF survey data. In FY2016, for instance, the Department of Defense reported $24.6 billion on procurement contract obligations for R&D services. In addition, the agency reported approximately $6 billion in grants (for all types of grants, including but not limited to R&D grants). Therefore, we expect to see a figure of under $31 billion for the DoD in the Federal Funds for R&D Survey data for FY2016.

[5]The bar graph plots the aggregated annual research and development expenditure by source of funds from the *NSF National Patterns of R&D Resources (2014-15)*, tables 8 and 9 (the "Other" category aggregates non-federal government, higher education, and other non-profits). The line "Federally Sourced, Performed by Business" is a subset of the federal spending on R&D that is performed by the business sector (the others performing sectors are federal intramural, FFRDC, non-federal government, higher education, and other non profits). The line "Federal R&D Procurement" plots the Federal R&D procurement dollars used in the construction of the instrumental variable in section 2.6.4. Figures are adjusted to 2012 dollars using GDP deflators from Louis Johnston and Samuel H. Williamson, "What Was the U.S. GDP Then?" MeasuringWorth, 2020.

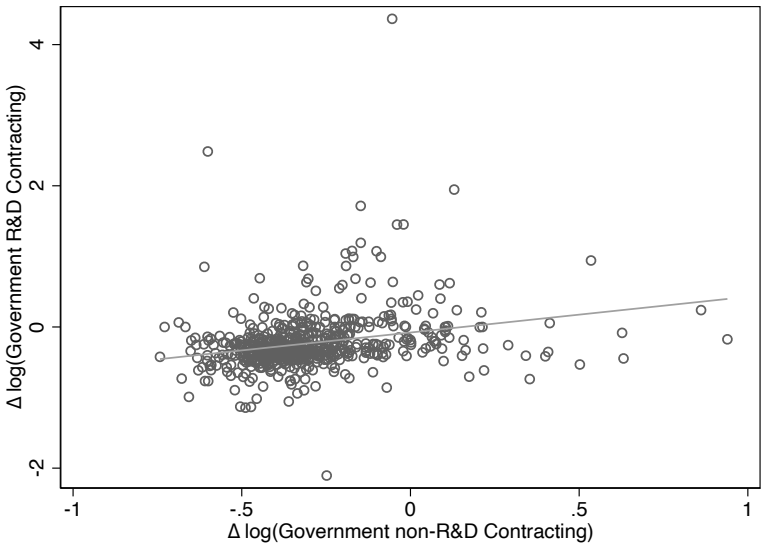


FIGURE A.3: Funding Shocks for R&D vs non-R&D Procurements[6]

However, the FFS reports $42 billion, which results in a difference of $11 billion. It is possible that some data was omitted because the contracted amounts were below the reporting thresholds (ranging between $2,000 and $10,000 depending on item specific requirements), or due to national security concerns. Pece (2016) also points out that discrepancies between accounting systems maintained by the respective agencies make collection of consistent data difficult.

Third, the NSF survey data includes grants, which we have not collected. The NSF "National Patterns" report relies on the Survey of Federal Funds for Research and Development (FFS) and the Census Bureau's Business R&D and Innovation Survey (BRDIS). The questionnaires in these surveys collect data on not only contracts but also grants.[7]

[6]This plots the difference in federal funding before (1986-88) and after (1990-92) the fall of the Berlin Wall at the 4 digit IPC for R&D funding (y-axis) against non-R&D funding (x-axis) (r=0.257).

[7]The 2001 questionnaire for the FFS states that "A performer [of R&D] is either an intramural group or organization carrying out an operational function or an extramural organization or person receiving support or providing services under a contract **or grant**." (brackets and emphasis added by authors)

We address the concern that this omission may systematically undercount certain scientific disciplines. For instance, the life sciences relies heavily on grants from the National Institutes of Health (NIH). We therefore compare the R&D procurement spending for the life sciences against the R&D outlays for the Department of Health and Human Services (under which the NIH is classified). The NSF data indicates a 35% increase in real dollar terms (from $9.9 to $13.5 billion 2012 dollars). Our data on the HHS also shows a 5.7 times increase for the same period, while R&D contracts for medical (2 digit PSC code: "AN") purposes increase 5.3 times. Therefore, the direction of the change is same, if the magnitudes are different.

A.1.3 IVE Robustness Check

We calculate a new instrument that calculates funding differences after controlling for differences in MFT demand conditions. This instrument predicts funding shocks net of controls for patenting quantity, quality, and propensity to trade from an OLS specification. We estimate the following OLS specification for 4-digit IPC k and = year t for years 1986 through 1992:

$$\begin{aligned} Government\ R\&D\ Contracting_{kt} = \beta_0 + \beta_1 Post1989_t \\ + \beta_2 Post1989_t \times IPC_Dummy_i \\ + IPC_Dummy_k + Year_Dummy_t + \boldsymbol{Z'_{kt}} + \ \epsilon_{kt} \end{aligned} \tag{A.8}$$

where controls $\boldsymbol{Z'_{it}}$ consist of number of patents, forward patent citations, and share of patents traded. $Government\ R\&D\ Contracting_{kt}$ is the amount of government R&D funding relevant to each 4-digit IPC-year.[8] For each 4-digit IPC, the logged difference in predicted government funding due to the end of the Cold

[8]The funding data is aggregated at the PSC-year level and crosswalked to the 4-digit IPC year level. The weights are based on the relevance of each 4-digit IPC to a PSC using patenting data of vendor names matched to the DISCERN database.

War net of the controls is $\Delta log(\widehat{Government\ R\&D\ Contracting}) := log(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 IPC_Dummy_i + IPC_Dummy_i) - log(\hat{\beta}_0 + IPC_Dummy_i)$. This is the instrument used in columns 3 and 4 of table A.3.

Table A.3: Post Cold War Federal R&D Shifts and MFT (Predicted Funding Shocks)

	(1)	(2)	(3)
	OLS	1st Stage IV	2nd Stage IV
Dependent Variable:	ln(Share of Reassigned Patents)	ln(Avg Cites to Science)	ln(Share of Reassigned Patents)
ln(Avg Cites to Science)	1.040**		0.650**
	(0.056)		(0.085)
Δ ln(Gov. R&D Contracting) (Predicted)		0.306**	
		(0.005)	
ln(Gov. R&D Contracting (Pre, $1Bn))	-0.476**	0.122**	-0.481**
	(0.086)	(0.016)	(0.081)
ln(Gov. non-R&D Contracting (Pre, $1Bn))	0.314**	0.008*	0.295**
	(0.031)	(0.004)	(0.032)
ln(Number of Patents)	-0.042**	-0.003	-0.027**
	(0.007)	(0.006)	(0.006)
Share of Small Assignees	0.396**	-0.276**	0.305**
	(0.052)	(0.004)	(0.054)
Avg of DV	2.059	0.109	2.059
SD of Science	0.128		0.128
Cragg-Donald F-Stat		321.527	
Year Fixed Effects	Yes		Yes
R^2	0.119		
N	1,928	1,928	1,928

Notes: Analysis is at the 4-digit IPC-patent publication year level. Sample period is 1992 and 2000 inclusively for all columns. $\Delta log(Gov.\ R\&D\ Contracting)(Predicted)$ is calculated as the logged difference in predicted values from an OLS specification in equation A.8. Other variable definitions are identical to table 2.10. All specifications include patent publication year fixed effects. Standard errors are clustered at the year level.

A potential threat to the validity of table 2.10 in section 2.6.4 is that our results are biased by secular patterns in MFT preceding our sample period between 1992 and 2000. For instance, it is possible that areas where federal R&D spending increased were also areas where patent market regulations were selectively relaxed or random scientific discoveries were concentrated in. If so, then we should expect there to be a considerable difference in the level of MFT activity for IPCs whose federal R&D

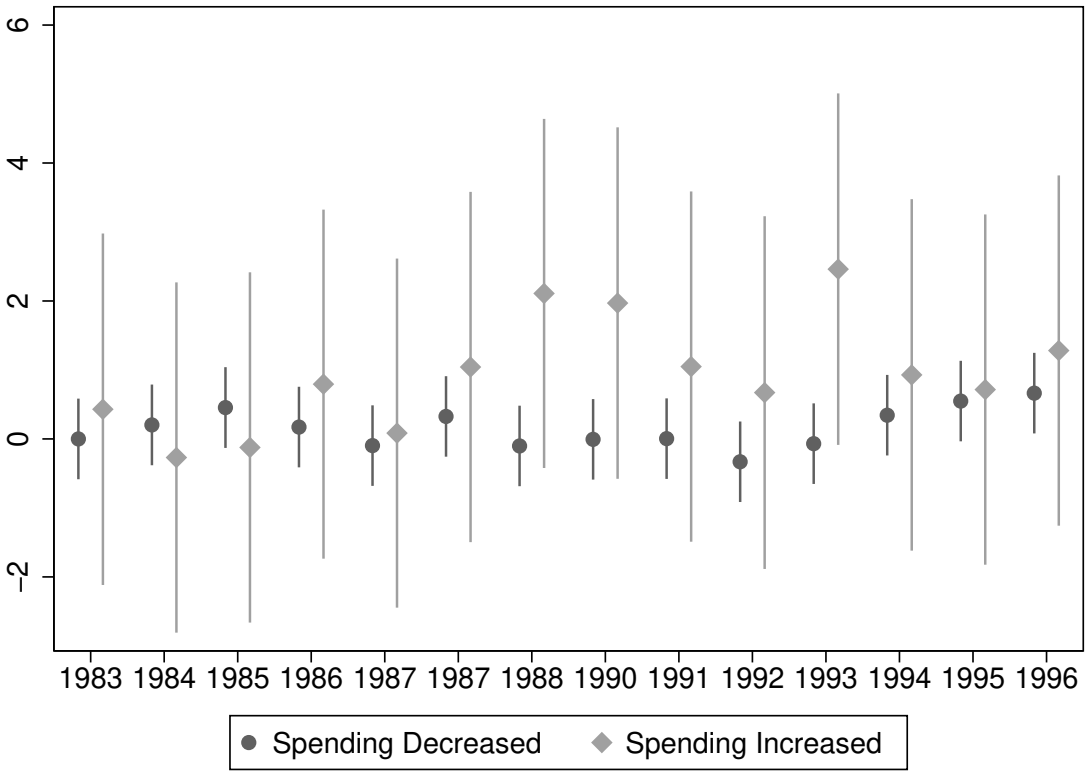


FIGURE A.4: Share of Patents Traded vs Shock[9]

spending increased and those whose spending decreased. We regress at the 4-digit IPC-year level the share of traded patents against year dummies after splitting the sample where average Federal R&D spending increased and those where spending decreased. Figure A.4 plots the coefficients of these OLS specifications (where 1982 is the base year) and finds that there is no statistically significant difference between the groups positively and negatively affected by federal spending shocks.

[9]This plots the coefficients from regressing share of patents traded within 5 years of grant at the 4 digit IPC-year level against patent publication years, splitting the sample by whether the focal IPC was negatively (dark circle marker) or positively (light diamond marker) shocked by the end of the Cold War. The base group is 1982.

We find in column 1 of table 2.10 that a standard deviation larger government contract R&D shock around the end of the Cold War leads to a 7% increase in papers in the relevant Web of Science field.

Given that corporate share in total scientific publications is generally low compared to academia,[10] the existence of the 7% effect may be suspect if one assumes that government contract R&D procurement is principally carried out by the corporate sector. Nevertheless, we find that between 1986 and 1992 (when the funding shocks are calculated), U.S. public firms account for only around 22% ($6.7 Bn) of all contract R&D (which averages $32 Bn) in our procurement data. Moreover, around a third (114/357) of R&D contracting firms never publish in science, while only around 40% operate a corporate lab for 1988 and 1991.[11] This suggests that a sizable share of the contract R&D work is sub-contracted to public entities that may in turn publish follow-on research.

This conjecture is consistent with prior research that argues that American firms started to "externalize" their R&D operations "through such mechanisms as consortia, collaboration with US universities and federal laboratories" since the beginning of the 1980s (Mowery, 1998, p.646). On the other hand, universities began to rely more heavily on industry support: between 1960 and 1995, industry contribution to university research tripled to 7%, while more than 1050 research institutes "seeking to support research on issues of direct interest to industry" were being run by 1992

[10]For example, the NSF's 2018 Science and Engineering Indicators produces well over 70% of all peer-reviewed scientific papers between 2003 and 2016 (https://www.nsf.gov/statistics/2018/nsb20181/report/sections/academic-research-and-development/outputs-of-s-e-research-publications)

[11]We link our procurement data to a comprehensive directory of all corporate industrial laboratories in the United States from Png (2019). The dataset is available at https://scholarbank.nus.edu.sg/handle/10635/150104 and contains the number of professionals, doctorates, and technicians reported by American firms for 1981, 1983, 1985, 1988, 1991, 1994, 1997. The dataset also links lab names to Compustat GVKEYs, which allows us to link the data to the procurement data.

(Mowery, 1998, p.648).[12]

A.2 Auxiliary Results

Table A.4: Reliance on Science and MFT, Forward Patent Citation Controls)

	DV: Pr(Reassignment=1)		
	(1) Continuous	(2) Quintiles	(3) Deciles
Cite Science Dummy	0.780**	0.825**	0.802**
	(0.039)	(0.038)	(0.038)
5-year Forward Patent Cites	0.104**		
	(0.002)		
Triadic Patent Dummy	0.810**	0.820**	0.809**
	(0.029)	(0.029)	(0.029)
Number of Claims	0.069**	0.070**	0.069**
	(0.001)	(0.001)	(0.001)
Length of First Claim	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
Forward Patent Citations (1st Quintile)		0.000	
		(.)	
Forward Patent Citations (2nd Quintile)		0.302**	
		(0.035)	
Forward Patent Citations (3rd Quintile)		0.673**	
		(0.036)	
Forward Patent Citations (4th Quintile)		1.293**	
		(0.037)	
Forward Patent Citations (5th Quintile)		2.593**	
		(0.040)	
Forward Patent Citations (1st Decile)			0.000
			(.)
Forward Patent Citations (2nd Decile)			-0.098
			(0.052)
Forward Patent Citations (3rd Decile)			0.175**
			(0.046)
Forward Patent Citations (4th Decile)			0.385**
			(0.046)
Forward Patent Citations (5th Decile)			0.546**
			(0.047)
Forward Patent Citations (6th Decile)			0.757**
			(0.048)
Forward Patent Citations (7th Decile)			1.131**
			(0.049)
Forward Patent Citations (8th Decile)			1.414**
			(0.050)
Forward Patent Citations (9th Decile)			1.900**
			(0.051)
Forward Patent Citations (10th Decile)			3.262**
			(0.055)
Avg of DV	6.265	6.265	6.265
4-digit IPC Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Firm Fixed Effects			
R^2	0.013	0.013	0.013
N	3,882,632	3,882,632	3,882,632

Notes: Unit of analysis is at the patent level. "5-year Forward Patent Cites" counts the number of forward patent citations a focal patent receives. Column 2 and 3 respectively include dummies for quintiles and deciles of "5-year Forward Patent Cites" by patent grant year and 4-digit IPC. Other variable definitions are identical to those in table 2.3

[12] 57% of these were established during the 1980s.

Table A.5: Post Cold War Federal R&D Shifts and MFT (Bootstrapped Standard Errors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	1st Stage IV	2nd Stage IV	1st Stage IV	2nd Stage IV	1st Stage IV	2nd Stage IV
Dependent Variable:	ln(Share of Reassigned Patents)	ln(Avg Cites to Science)	ln(Share of Reassigned Patents)	ln(Avg Cites to Science)	ln(Share of Reassigned Patents	ln(Avg Cites to Science)	ln(Share of Reassigned Patents
ln(Avg Cites to Science)	1.040**		0.905**		0.666**		0.650**
	(0.051)		(0.093)		(0.201)		(0.197)
Number of Papers (Predicted, 1000s)		0.075**					
		(0.003)					
Δ ln(Gov. R&D Contracting)				0.189**			
				(0.009)			
Δ ln(Gov. R&D Contracting) (Predicted)						0.316**	
						(0.016)	
ln(Gov. R&D Contracting (Pre, $1Bn))	-0.476**	-0.005	-0.478**	0.125**	-0.481**	0.094**	-0.481**
	(0.078)	(0.029)	(0.091)	(0.030)	(0.092)	(0.029)	(0.088)
ln(Gov. non-R&D Contracting (Pre, $1Bn))	0.314**	-0.062**	0.307**	-0.037**	0.296**	0.034**	0.295**
	(0.028)	(0.012)	(0.043)	(0.012)	(0.043)	(0.012)	(0.042)
ln(Number of Patents)	-0.042**	-0.002	-0.037*	0.022**	-0.027	-0.010*	-0.027
	(0.007)	(0.004)	(0.016)	(0.004)	(0.017)	(0.004)	(0.016)
Share of Small Assignees	0.396**	-0.244**	0.364**	-0.242**	0.308**	-0.263**	0.305**
	(0.047)	(0.011)	(0.053)	(0.012)	(0.068)	(0.012)	(0.064)
Avg of DV	2.059	0.109	2.059	0.109	2.059	0.109	2.059
SD of Science	0.128		0.128		0.128		0.128
Cragg-Donald F-Stat		700.301		358.905		320.356	
Year Fixed Effects	Yes		Yes		Yes		Yes
R^2	0.119						
N	1,928	1,928	1,928	1,928	1,928	1,928	1,928

Notes: This table replicates the results in table 2.10 and A.3 with bootstrapped standard errors. Analysis is at the 4-digit IPC-patent publication year level. All specifications include patent publication year fixed effects. Standard errors are from 1000 bootstrapped samples.

Appendix B

Appendix to Chapter 3

B.1 Model

B.1.1 Setup

There are three stages. In stage 3, the firms compete in the product market. Their product market performance depends on the quality of their products and the cost of producing them. We assume that cost and quality depend upon the innovation output, d_i, $i = 0, 1$. Their payoffs from stage 3 are $\Pi(d_0, d_1)$ and $\tilde{\Pi}(d_1, d_0)$, where the tilde indicates firm 1. We assume that $\Pi(d_0, d_1)$ is increasing in the first argument and decreasing in the second, and concave in its arguments, so that the firm's profit increases in its innovation output, albeit at a diminishing rate. To avoid the need for assumptions on third order derivatives, we assume

$$\Pi(d_0, d_1) = kd_0 - \frac{c_{00}}{2}d_0^2 - bd_1 - \frac{c_{11}}{2}d_1^2 + c_{01}d_1d_0, \quad k > 1$$

$$\tilde{\Pi}(d_1, d_0) = d_1 - \frac{c_{00}}{2}d_1^2 - bd_0 - \frac{c_{11}}{2}d_0^2 + c_{01}d_1d_0$$

Firms farther from the frontier (e.g., smaller firms) can increase profits by imitation and by increasing scale, possibilities that the leaders have already exhausted. Instead, leaders have to introduce new and improved products and processes-to innovate. Accordingly, the marginal product of innovation for firm 0 is greater than that of firm 1 because $k > 1$.

The coefficient c_{01} is positive under strategic complementarity and negative under substitutability. Concavity of Π implies $c_{00} > 0, c_{11} > 0, c_{00}c_{11} - c_{01}^2 \geqslant 0$. We assume that $b > 0$ so that $\frac{\partial \Pi}{\partial d_1} = -b - c_{11}d_1 < 0$, i.e., innovation by rivals reduces payoff. We also assume that $c_{00} \geqslant c_{11}$. This assumption implies that the marginal returns to internal invention decline faster than the rate at which profits decline due to invention by rivals.

In stage 2, firms choose their innovation output. Firm 0 chooses d_0 and firm 1 chooses d_1. The cost of innovation for firm 0 is $\phi(r_0; u)d_0$, where r_0 represents investments in internal scientific research by the firm, and u indexes the stock of (relevant) public science. The cost of innovation includes the cost of inventing new products and processes or improving them. Internal research may directly lead to such inventions, but may also indirectly reduce the cost of invention by guiding the search for inventions in more promising directions. Innovations may also be based on inventions acquired from independent inventors, other firms or university researchers. Thus the cost of innovation also depends on the state of public science. It is natural to assume that both internal research and public science reduce the unit cost of innovation, $\phi(r_0; u)$, i.e., $\frac{\partial \phi}{\partial r_0} < 0$, $\frac{\partial \phi}{\partial u} < 0$, and diminishing returns so that $\frac{\partial^2 \phi}{\partial r_0^2} > 0$.

As we show below, the relationship between public science and internal research in the reduction in the unit cost of innovation will be important in how research investments relate to the stock of public science. The relationship may be one of strategic complementarity (in the sense of Milgrom and Roberts 1989). For instance, it is typically believed that public science would complement internal research efforts. However, public science may also lead to startups and independent inventors, who can license or sell their inventions, which can substitute for internally generated inventions. If so, the relationship may be one of strategic substitutability. Strategic complementarity exists if $-\frac{\partial^2 \phi}{\partial r_0 \partial u} > 0$, and substitutability exists if $-\frac{\partial^2 \phi}{\partial r_0 \partial u} < 0$. If $\frac{\partial^2 \phi}{\partial r_0 \partial u} = 0$, public science and research have independent effects on the cost of innovation.

The cost of innovation for firm 1 is $\phi(\tilde{u})d_1$. As noted, innovations may be based on external discoveries and inventions. Thus, we assume that $\phi(\tilde{u})$ decreases with u.

In stage 1, firm 0 choose its research investments, r_0, and the cost of research is modelled simply as $\frac{\gamma}{2} r_0^2$, so $v_0 = d_0 - \frac{c_{00}}{2} d_0^2 - b d_1 - \frac{c_{11}}{2} d_1^2 + c_{01} d_1 d_0 - \phi(r_0, \lambda) d_0 - \frac{\gamma}{2} r_0^2$.

B.1.2 Stage 2: Innovation

We assume a stable Nash Equilibrium exists. For a stable equilibrium, we require that $D = c_{00}^2 - c_{01}^2 > 0 \iff |c_{00}| > |c_{01}|$.

Note that as long as $k \geqslant 1 + (\phi - \tilde{\phi})$, $d_0 \geqslant d_1$. In particular, if neither firm invests in research, so that $\phi = \tilde{\phi}$, firm 0 would innovate more, and the gap increases the larger is k. This would imply that firm 0 has a greater incentive to invest in research. The following intermediate results are helpful for later results.

Focal Firm Research and Innovation

The response of innovation output to the focal firm's research is

$$\begin{aligned} \frac{\partial d_0}{\partial r_0} &= \frac{c_{00}}{D}(-\frac{\partial \phi}{\partial r_0}) \\ \frac{\partial d_1}{\partial r_0} &= \frac{c_{01}}{D}(-\frac{\partial \phi}{\partial r_0}) \end{aligned} \tag{B.1}$$

Note that if $c_{01} \geqslant 0$, firm 1 also increases its innovation in response to an increase in research by firm 0. Furthermore, $\frac{\partial^2 d_0}{\partial r_0 \partial u} = -\frac{c_{00}}{D}\frac{\partial^2 \phi}{\partial r_0 \partial u} \geqslant 0$ if $\frac{\partial^2 \phi}{\partial r_0 \partial u} \leqslant 0$, i.e., if public science and internal research are complements.

Public Science and Innovation

The response of innovation output to public science is

$$\begin{aligned} \frac{\partial d_0}{\partial u} &= \frac{-1}{D}\left(c_{00}\frac{\partial \phi}{\partial u} + c_{01}\frac{\partial \tilde{\phi}}{\partial u}\right) \\ \frac{\partial d_1}{\partial u} &= \frac{-1}{D}\left(c_{00}\frac{\partial \tilde{\phi}}{\partial u} + c_{01}\frac{\partial \phi}{\partial u}\right) \end{aligned} \tag{B.2}$$

If there is strategic complementarity, i.e., $c_{01} \geqslant 0$, both firms innovate more in response to an increase in public science. However, if there is strategic substitutability, then one (but not both) firms may reduce innovation. In particular, if the innovation

costs of a firm are not very responsive to public science, the effect of a rival increasing its innovation may cause the firm to reduce its innovation.

B.1.3 Stage 1: Research

Suppose firm 1 does not invest in research. Firm 0 chooses r_0, taking into account how its choice will affect the equilibrium choices of d_0 and d_1 in the stage 2 game. For firm 0, the first-order condition for optimal r_0, is

$$-\frac{\partial \phi}{\partial r_0} d_0 + \frac{\partial \Pi}{\partial d_1}\frac{\partial d_1}{\partial r_0} = \gamma r_0 \tag{B.3}$$

The marginal return to research has a direct benefit represented by the first term: the reduction in the unit cost of innovation, which is proportional to the scale of innovation. The second term represents the feedback effect from competition in the innovation stage. By increasing innovation, research has a secondary benefit if it reduces innovation by the rival, which would be the case if there is strategic substitution in the innovation, so that $c_{01} \leqslant 0$. If innovations are strategic complements, then there is a secondary cost, because the second term would be negative. However, the first term is always larger than the second term. Substituting for $\frac{\partial d_1}{\partial r_0}$ from Equation B.1 and gathering terms, Equation B.3 can be rewritten as

$$-\frac{\partial \phi}{\partial r_0}\left(\frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D} + d_0\right) = \gamma r_0 \tag{B.4}$$

Therefore, $\frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D} + d_0$ must be positive. A sufficient condition for this is strategic substitutability in innovation, $c_{01} \leqslant 0$. [1]

[1] We assume that the second order condition for an interior maximum holds. This requires that γ be large.

B.1.4 Result 1: Innovation Leadership

Leaders earn higher profits. Conversely, the profits of the follower fall with the lead of firm 0. Formally,

$$\begin{aligned}\frac{\partial v}{\partial k} &= d_0 + \frac{\partial \Pi}{\partial d_1}\frac{\partial d_1}{\partial k} \\ &= d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D} > 0 \text{ at an interior maximum} \\ \frac{\partial \tilde{v}}{\partial k} &= \frac{\partial \tilde{\Pi}}{\partial d_0}\frac{\partial d_0}{\partial k} = \frac{\partial \tilde{\Pi}}{\partial d_0}\frac{c_{00}}{D} < 0\end{aligned} \tag{B.5}$$

Importantly, the returns to research of the innovation leader increase with its lead k. Those of the follower decrease if innovations are strategic substitutes and increase otherwise. Intuitively, as k increases, the leader increases innovation. With strategic substitutes, the marginal return to innovation for the follower decreases. Given that research reduces the cost of innovation, the marginal return to research for the follower decreases.

$$\begin{aligned}\frac{\partial^2 v}{\partial k \partial r_0} &= \frac{\partial d_0}{\partial r_0} + \frac{c_{01}}{D}(-c_{11}\frac{\partial d_1}{\partial r_0} + c_{00}\frac{\partial d_0}{\partial r_0}) \\ &= \frac{c_{00}}{D}(-\frac{\partial \phi}{\partial r_0}) + (-\frac{\partial \phi}{\partial r_0})\frac{c_{01}^2}{D}(c_{00} - c_{11}) > 0 \\ \frac{\partial^2 \tilde{v}}{\partial k \partial r_1} &= \frac{c_{00}}{D}(-c_{11}\frac{\partial d_0}{\partial r_0} + c_{01}\frac{\partial d_1}{\partial r_1}) = \frac{c_{00}}{D}(-\frac{\partial \tilde{\phi}}{\partial r_1})c_{01}(c_{00} - c_{11}) \leqslant 0 \iff c_{01} \leqslant 0\end{aligned} \tag{B.6}$$

This result points to why firm 1 may not invest in research. If k is large and there is strategic substitutability, firm 1's scale of innovation is small, thereby reducing its returns to innovation.

B.1.5 Public Science

Equation B.6 implies that if innovations are strategic substitutes, as the gap between leaders and followers grows, their incentives to invest in research diverge: leaders are

more likely to invest in research, and followers are less likely to do so. If there is a fixed cost to such investment, then, for a range of such costs, we will have only firm 0 invest in research while firm 1 does not. In this section, we focus on the equilibrium where only firm 0 invests in research.

The Value of the Firm

The value of the firm, v, may decrease with public science if public science substitutes for internal research, particularly if innovations are strategic complements. Intuitively, although public science reduces the cost of innovation, the innovation cost of the rival also declines. Increased innovation by the rival reduces value for the focal firm. If public science substitutes for internal research, it will be less effective in reducing the innovation cost of firm 0, i.e., $|\frac{\partial \phi}{\partial u}| < |\frac{\partial \tilde{\phi}}{\partial u}|$. Formally, the value of the firm is $v = \max_{r_0}\{\Pi - \gamma \frac{r_0^2}{2}\}$. Applying the envelope theorem, the effect of public science is given by

$$\begin{aligned} \frac{\partial v}{\partial u} &= -d_0 \frac{\partial \phi}{\partial u} + \frac{\partial \Pi}{\partial d_1}\frac{\partial d_1}{\partial u} \\ &= -\frac{\partial \phi}{\partial u}\left(d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D}\right) - c_{00}\frac{\partial \Pi}{\partial d_1}\frac{\partial \tilde{\phi}}{\partial u} \end{aligned} \tag{B.7}$$

Although the first term is positive by Equation B.3, its magnitude depends on $|\frac{\partial \phi}{\partial u}|$. The second term is negative, and represents the effect due to the reduction in the rival's innovation cost. It is larger in magnitude the larger is $|\frac{\partial \tilde{\phi}}{\partial u}|$. Note that rivalry also matters. If $\frac{\partial \Pi}{\partial d_1} = -b + c_{01}d_0$ is large in magnitude (as would be the case for b large and $c_{01} < 0$, the firm's value can decline with public science.

B.1.6 Internal Research and Public Science

At an interior maximum, the direction of the effect of public science on internal research is given by $\frac{\partial^2 v}{\partial r_0 \partial u}$. Research increases with public science if $\frac{\partial^2 v}{\partial r_0 \partial u} \geqslant 0$ and decreases otherwise.

$$\frac{\partial^2 v}{\partial r_0 \partial u} = \left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} + d_0\left(-\frac{\partial^2 \phi}{\partial r_0 \partial u}\right) + \frac{\partial \Pi}{\partial d_1}\frac{\partial^2 d_1}{\partial r_0 \partial u} + \frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u}$$

substituting and collecting terms (B.8)

$$= \left(-\frac{\partial \phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} - \frac{\partial^2 \phi}{\partial r_0 \partial u}\left(d_0 + \frac{\partial \Pi}{\partial d_1}\frac{c_{01}}{D}\right) + \frac{\partial d_1}{\partial r_0}\frac{\partial^2 \Pi}{\partial d_1 \partial u}$$

The first term in Equation B.8 is positive. The second is positive if public science and research are strategic complements and negative otherwise. The third term is negative only if innovations are strategic complements and positive otherwise. Put differently, the first term reflects a direct effect: public science reduces innovation costs, and the resulting increase in innovation increases the marginal return to research. The second term represents the interaction between public science and research in reducing innovation costs. If they are complements, the second term also implies that the marginal return to research increases with public science. The third term captures the strategic interaction in innovation. If innovations are strategic substitutes, this term is also positive. Strategic complementarity is a necessary, but not sufficient, condition for this term to be negative. Thus, if internal research falls with public science, it implies that public science is a strategic substitute for research, or innovations are strategic complements, or both. These are one-way implications; even if they hold, public science could increase internal research if the direct effect, represented by the first term, is large.

To see this more fully, consider the case where there is neither complementarity

nor substitution in the innovation stage, and where public science and research are independent. The latter implies that $\frac{\partial^2\phi}{\partial r_0 \partial u} = 0$, and the former implies that $\frac{\partial d_1}{\partial r_0} = 0$. In that case, Equation B.8 has a single term $\left(-\frac{\partial\phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} \geqslant 0$. *That is, if public science and research are independent and there are no strategic interactions in the innovation stage, internal research increases with public science because public science increases the scale of innovation, thereby increasing the marginal return to research.*

If there are no strategic interactions in innovation, Equation B.8 is $\left(-\frac{\partial\phi}{\partial r_0}\right)\frac{\partial d_0}{\partial u} - \frac{\partial^2\phi}{\partial r_0 \partial u}\left(d_0 + \frac{\partial\Pi}{\partial d_1}\frac{c_{01}}{D}\right)$. The second term is non-negative if $-\frac{\partial^2\phi}{\partial r_0 \partial u} \geqslant 0$, i.e., if public science and internal research are complements and negative otherwise. *Therefore, if internal research declines with public science, and there are not strategic interactions in innovation, it implies that public science and internal research are strategic substitutes.*

The third term can be written as

$$\frac{\partial d_1}{\partial r_0}\frac{\partial^2\Pi}{\partial d_1 \partial u} = \frac{\partial d_1}{\partial r_0}\left[\frac{\partial^2\Pi}{\partial d_1 \partial d_0}\frac{\partial d_0}{\partial u} + \frac{\partial^2\Pi}{\partial d_1^2}\frac{\partial d_2}{\partial u}\right]$$

$$= \frac{\partial d_1}{\partial r_0}\frac{1}{D}\left[-c_{11}c_{00}(-\frac{\partial\tilde{\phi}}{\partial u}) + -c_{11}c_{01}(-\frac{\partial\phi}{\partial u}) + c_{01}c_{00}(-\frac{\partial\phi}{\partial u}) + c_{01}^2(-\frac{\partial\tilde{\phi}}{\partial u})\right]$$

collecting terms and substituting

$$\frac{\partial d_1}{\partial r_0}\frac{\partial^2\Pi}{\partial d_1 \partial u} = \frac{c_{01}}{D^2}(-\frac{\partial\tilde{\phi}}{\partial u})(c_{01}^2 - c_{00}c_{11}) + \frac{c_{01}^2}{D^2}(-\frac{\partial\phi}{\partial u})(c_{00} - c_{11}) \tag{B.9}$$

Note that $c_{00} \geqslant c_{11}$, so that $\frac{c_{01}^2}{D^2}(-\frac{\partial\phi}{\partial u})(c_{00} - c_{11}) \geqslant 0$. Also, $-\frac{\partial\tilde{\phi}}{\partial u}(c_{01}^2 - c_{00}c_{11}) \leqslant 0$ by the concavity of Π. Thus, $\frac{c_{01}}{D^2}(-\frac{\partial\tilde{\phi}}{\partial u})(c_{01}^2 - c_{00}c_{11}) > 0$ if $c_{01} < 0$ and negative otherwise. Therefore, a necessary condition for the expression in Equation B.9 to be negative is that innovations be strategic complements. The conclusion is that for public science to reduce research, it would require that either innovations be strategic

complements, or that public science be a strategic substitute for internal research. Else, public science will increase research by the leading firm.

The Gap Between the Leader and Follower, the Returns to Research, and Public Science

Recall from Equation B.6 that the marginal returns from research to the leader as k increases is given by $\frac{c_{00}}{D}(-\frac{\partial\phi}{\partial r_0}) + (-\frac{\partial\phi}{\partial r_0})\frac{c_{01}^2}{D}(c_{00} - c_{11})$. It is easy to see that this expression is increasing in u if public science and internal research are complements – $\frac{\partial^2\phi}{\partial r_0 \partial u} \geqslant 0$–and decreasing otherwise. Similarly, the effect on the marginal returns from research to the follower of k increases is given by $\frac{c_{00}}{D}(-\frac{\partial\tilde{\phi}}{\partial r_1})c_{01}(c_{00} - c_{11})$. This expression falls with u if innovations are strategic complements, $c_{01} > 0$, and public science and research are substitutes, $\frac{\partial^2\tilde{\phi}}{\partial r_1 \partial u} \geqslant 0$, or if innovations are strategic substitutes but public science and research are complements. Otherwise, the marginal returns of the follower also increase with u.

B.2 Auxiliary Results

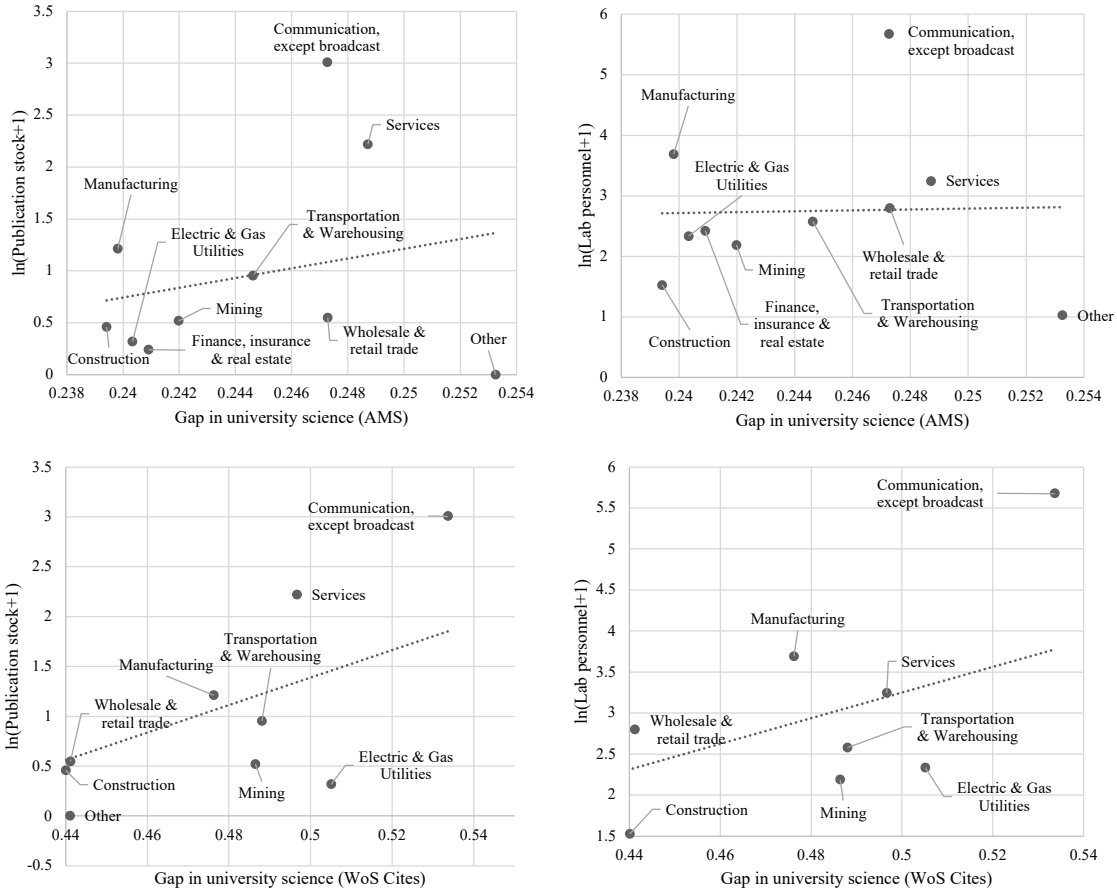


FIGURE B.1: Corporate Science vs Gaps in University Science, by Industry[2]

In Table B.1, we replicate the result in Tables 3.4 and 3.5 by using the AMS scientist-based measure of the gap. The directions of the interaction term coefficient for forward patent citations is identical, though we fail to replicate statistical significance. However, the coefficient for patent citations to science is positive and significant. The effects of corporate size, group affiliation and competitive industries

[2]Industry-level scatter plots of firm investment in science and the gaps in the relevant academic discipline. The left panels plot the natural log of one plus the average publication stock against the gaps in public science measures. The right panel replaces the publication stock with the number of personnel at R&D labs from the IRL directory. The upper panels measure gaps using the number of foreign-trained American scientists in the AMS directory. The lower panels measure gaps using share of transatlantic citations made by American journals.

all have the same signs and are statistically significant, as in Tables 3.4 and 3.5. Table B.2 uses the journal citations-based gap measure. All results except the interaction with patent citations to science (Column 2) are also statistically significant at the 5% level (in Column 2, the interaction coefficient is significant at the 10% level (t= 1.95)).

Table B.1: Corporate Publications and Gaps in University Science (Scientists), by Firm Characteristics

	DV: Publication stock					
	(1)	(2)	(3)	(4)	(5)	(6)
Gap in university science, 1921 (Scientists) × Patenting Intensity	183.739					
	(103.126)					
Gap in university science, 1921 (Scientists) × Forward Patent Citations		129.790				
		(40.331)				
Gap in university science, 1921 (Scientists) × ln(Patent Cites to Science)			4171.920			
			(792.562)			
Gap in university science, 1921 (Scientists) × ln(Assets) (1926-1930)				439.018		
				(71.239)		
Gap in university science, 1921 (Scientists) × Business Group Dummy (Imputed)					623.660	
					(243.421)	
Gap in university science, 1921 (Scientists) × Competitive market						-447.074
						(67.875)
Gap in university science, 1921 (Scientists)	-21.778	193.082	213.533	-11315.717	233.304	420.305
	(40.215)	(65.536)	(61.361)	(1840.963)	(167.575)	(83.332)
Forward Patent Citations		-29.004				
		(9.543)				
ln(Patent Cites to Science)			-981.860			
			(197.172)			
ln(Assets) (1926-1930)				-101.985		
				(16.710)		
Business Group Dummy (Imputed)					-152.217	
					(58.545)	
Patenting Intensity	-38.056					
	(25.508)					
ln(Assets)	-0.668	4.062	1.231	0.817	6.478	4.110
	(0.280)	(0.570)	(0.228)	(0.619)	(1.052)	(0.577)
ln(American scientists), 1921	8.372	16.108	21.474	15.549	18.959	11.425
	(4.388)	(7.211)	(7.082)	(7.360)	(18.070)	(7.135)
ln(European scientists), 1921	-7.155	-13.294	-18.677	-12.742	-10.873	-8.497
	(4.432)	(7.141)	(6.989)	(7.302)	(17.976)	(7.088)
ln(Patent Granted)			4.255			
			(0.672)			
Average of Dependent Variable	3.642	3.642	3.642	3.843	6.625	3.657
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.666	0.151	0.317	0.180	0.184	0.151
Number of Firms	422	422	422	358	199	420
Number of Observations	4,293	4,293	4,293	4,028	1,702	4,275

Notes: Unit of analysis is at the firm-year level. "Gap in university science, 1921 (Scientists)" is defined as the share of American scientists with a European affiliation (winsorized at 1% and 99%). The other variable definitions are identical to those in Tables 3.2, 3.4 and 3.5. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B.2: Corporate Publications and Gaps in University Science (Cites), by Firm Characteristics

	DV: Publication stock					
	(1)	(2)	(3)	(4)	(5)	(6)
Gap in university science, 1900-20 (Cites) × Patenting Intensity	39.254					
	(7.780)					
Gap in university science, 1900-20 (Cites) × Forward Patent Citations		16.649				
		(4.808)				
Gap in university science, 1900-20 (Cites) × ln(Patent Cites to Science)			508.747			
			(260.366)			
Gap in university science, 1900-20 (Cites) × ln(Assets) (1926-1930)				25.936		
				(4.890)		
Gap in university science, 1900-20 (Cites) × Business Group Dummy (Imputed)					36.773	
					(17.107)	
Gap in university science, 1900-20 (Cites) × Competitive market						-13.994
						(5.512)
Gap in university science, 1900-20 (Cites)	29.832	58.684	52.539	-579.706	183.879	77.714
	(5.756)	(13.663)	(10.589)	(117.909)	(32.884)	(15.127)
Patenting Intensity	-12.383					
	(3.627)					
Forward Patent Citations		-6.302				
		(2.181)				
ln(Patent Cites to Science)			-187.051			
			(122.358)			
ln(Assets) (1926-1930)				-8.169		
				(1.903)		
Business Group Dummy (Imputed)					-16.687	
					(7.683)	
ln(Assets)	-0.440	4.170	2.900	0.565	7.341	4.264
	(0.224)	(0.585)	(0.426)	(0.602)	(1.191)	(0.595)
ln(Cites to America), 1900-20	7.434	10.426	7.596	20.114	33.367	12.700
	(1.229)	(3.438)	(2.006)	(4.022)	(6.924)	(3.473)
ln(Cites to Europe), 1900-20	-6.433	-7.276	-4.938	-16.569	-25.661	-9.625
	(1.206)	(3.278)	(1.777)	(3.710)	(6.380)	(3.275)
Average of Dependent Variable	3.642	3.642	3.642	3.843	6.625	3.657
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.683	0.147	0.273	0.159	0.177	0.146
Number of Firms	422	422	422	358	199	420
Number of Observations	4,293	4,293	4,293	4,028	1,702	4,275

Notes: Unit of analysis is at the firm-year level. "Gap in university science, 1900-20 (Cites)" is defined at the firm level as American journal backward citations to Europe divided by total American backward journal citations (winsorized at 1% and 99%). The other variable definitions are identical to those in Tables 3.2, 3.4 and 3.5. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B.3: Corporate Science and "Home-Run" Patents (AMS Gaps)

	DV: Top 5% Xi			DV: Top 5% Novelty		
	(1) OLS	(2) 1st Stage IV	(3) 2nd Stage IV	(4) OLS	(5) 1st Stage IV	(6) 2nd Stage IV
Publication Stock (100s)	0.040		0.182	0.273		0.557
	(0.012)		(0.041)	(0.020)		(0.075)
Gap in university science, 1921 (Scientists)		250.154			202.045	
		(49.513)			(42.007)	
Gap in university science, current (Pubs)	-2.728	-2.779	-1.410	0.974	0.018	2.252
	(1.103)	(5.298)	(1.047)	(1.627)	(4.155)	(1.744)
ln(Assets)	1.419	2.383	1.080	0.761	2.273	0.110
	(0.218)	(0.458)	(0.186)	(0.160)	(0.384)	(0.210)
ln(Patent stock)	0.589	5.172	-0.139	1.866	4.170	0.689
	(0.081)	(0.559)	(0.144)	(0.152)	(0.473)	(0.239)
Average of Dependent Variable	0.864		0.861	2.903		2.895
Kleibergen-Paap rk Wald F statistic		25.525			23.134	
Year Fixed Effects	Yes		Yes	Yes		Yes
Industry Fixed Effects	Yes		Yes	Yes		Yes
R^2	0.248		-0.177	0.635		0.207
Number of Firms	327		327	425		425
Number of Observations	3,569	3,569	3,569	4,293	4,293	4,293

Notes: Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan, Papanikolaou, Seru and Stoffman, 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5, 6 is the share of European-affiliated scientists ("Gap in university science, 1921 (Scientists)"). "Gap in university science, current (Pubs)" calculates the share of European papers for the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B.4: Corporate Science and "Home-Run" Patents (WOS Citation Gaps)

	DV: Top 5% Xi			DV: Top 5% Novelty		
	(1) OLS	(2) 1st Stage IV	(3) 2nd Stage IV	(4) OLS	(5) 1st Stage IV	(6) 2nd Stage IV
Publication Stock (100s)	0.041		0.108	0.272		0.268
	(0.012)		(0.034)	(0.020)		(0.094)
Gap in university science, 1900-20 (Cites)		20.698			10.927	
		(4.848)			(3.642)	
Gap in university science, current (Cites)	-3.087	11.028	-4.774	1.788	14.544	1.874
	(1.040)	(4.910)	(1.356)	(1.082)	(3.996)	(2.282)
ln(Assets)	1.428	2.327	1.274	0.758	2.279	0.767
	(0.220)	(0.451)	(0.222)	(0.161)	(0.386)	(0.233)
ln(Patent stock)	0.582	5.261	0.232	1.871	4.238	1.888
	(0.080)	(0.568)	(0.167)	(0.153)	(0.482)	(0.411)
Average of Dependent Variable	0.864		0.861	2.903		2.895
Kleibergen-Paap rk Wald F statistic		18.225			9.001	
Year Fixed Effects	Yes		Yes	Yes		Yes
Industry Fixed Effects	Yes		Yes	Yes		Yes
R^2	0.248		0.112	0.636		0.592
Number of Firms	327		327	425		425
Number of Observations	3,569	3,569	3,569	4,293	4,293	4,293

Notes: Analysis is at the firm-year level. The dependent variable for Columns 1 and 3 is the number of firm patents in the top 5% of stock market value (Kogan, Papanikolaou, Seru and Stoffman, 2017). The dependent variable for Columns 4, 6 is the number of firm patents in the top 5% of novelty scores (Fleming, 2001). Columns 2 and 5 present first stage estimation results where dependent variable is publication stock. Instrument for Columns 2, 3, 5, 6 is the share of American journal citations made to European journals ("Gap in university science, 1900-20 (Cites)"). "Gap in university science, current (Cites)" calculates the backward citation share of European journals in American journals published in the focal year. Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

Table B.5: Corporate Science and Market-to-Book Ratios

	Cite Gap (OLS)		Scientist Gap (OLS)	
	(1)	(2)	(3)	(4)
	Small	Large	Small	Large
ln(Pubstock$_{t-1}$)	0.041	0.067	0.016	0.052
	(0.014)	(0.011)	(0.019)	(0.009)
ln(Patstock$_{t-1}$)	0.051	-0.019	0.026	0.009
	(0.006)	(0.006)	(0.008)	(0.005)
Gap in university science, current (Cites)	-0.332	0.175	-0.233	-0.101
	(0.185)	(0.185)	(0.163)	(0.121)
Average of Dependent Variable	0.617	0.563	0.605	0.580
Year Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.402	0.429	0.394	0.402
Number of Firms	169	156	149	176
Number of Observations	1,797	1,601	1,532	1,867

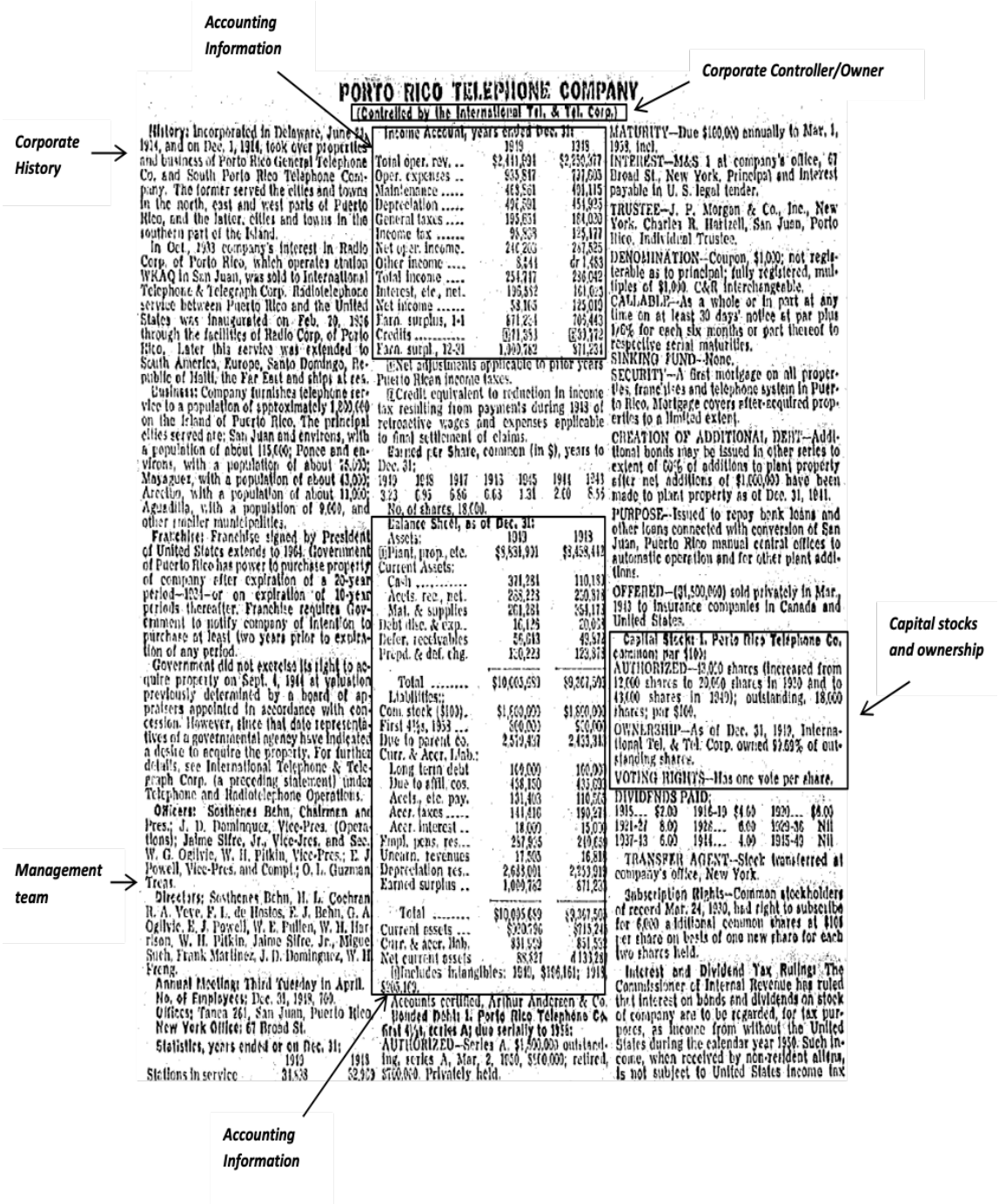
Notes: Unit of analysis is at the firm-year level. Dependent variable is Tobin's Q. Columns 1 and 2 are split by mean values of "Gap in university science, 1900-20 (Cites)" measures (1 being below average and 2 being above). Columns 3 and 4 are split by mean values of the "Gap in university science, 1921 (Scientists)" measures (3 being below average and 4 being above). Industry fixed effects are applied at 2-digit SIC codes. Standard errors are robust to arbitrary heteroscedasticity.

We replicate the same results of Table 3.7 in Columns 1 and 2 by splitting the sample by average citation gap scores. We find that the correlation between Q and gaps are 1.6 times larger for those above average gap scores compared to those below the average. A similar mean-split based on average AMS gaps (based on scientist bios) in Columns 3 and 4 shows the coefficients to be 3.3 times for higher gap firms.

B.3 Details on Data Construction

B.3.1 Corporate Historical Documents and Data Sources

- Bureau of Economic Analysis (BEA, 1958), U.S. Department of Commerce, Benchmark Federal Trade Commission (FTC) Annual Reports: www.ftc.gov/



PORTO RICO TELEPHONE COMPANY

(Controlled by the International Tel. & Tel. Corp.)

History: Incorporated in Delaware, June 23, 1914, and on Dec. 1, 1914, took over properties and business of Porto Rico General Telephone Co. and South Porto Rico Telephone Company. The former served the cities and towns in the north, east and west parts of Puerto Rico, and the latter, cities and towns in the southern part of the Island.

In Oct., 1933 company's interest in Radio Corp. of Porto Rico, which operates station WKAQ in San Juan, was sold to International Telephone & Telegraph Corp. Radiotelephone service between Puerto Rico and the United States was inaugurated on Feb. 20, 1936 through the facilities of Radio Corp. of Porto Rico. Later this service was extended to South America, Europe, Santo Domingo, Republic of Haiti, the Far East and ships at sea.

Business: Company furnishes telephone service to a population of approximately 1,800,000 on the Island of Puerto Rico. The principal cities served are: San Juan and environs, with a population of about 115,000; Ponce and environs, with a population of about 75,000; Mayaguez, with a population of about 43,000; Arecibo, with a population of about 11,000; Aguadilla, with a population of 9,000, and other smaller municipalities.

Franchise: Franchise signed by President of United States extends to 1964. Government of Puerto Rico has power to purchase property of company after expiration of a 20-year period—1924—or on expiration of 10-year periods thereafter. Franchise requires Government to notify company of intention to purchase at least two years prior to expiration of any period.

Government did not exercise its right to acquire property on Sept. 4, 1944 at valuation previously determined by a board of appraisers appointed in accordance with concession. However, since that date representatives of a governmental agency have indicated a desire to acquire the property. For further details, see International Telephone & Telegraph Corp. (a preceding statement) under Telephone and Radiotelephone Operations.

Officers: Sosthenes Behn, Chairman and Pres.; J. D. Dominguez, Vice-Pres. (Operations); Jaime Sifre, Jr., Vice-Pres. and Sec. W. G. Ogilvie, W. H. Pitkin, Vice-Pres.; E. J. Powell, Vice-Pres. and Compt.; O. L. Guzman Treas.

Directors: Sosthenes Behn, H. L. Cochran, R. A. Veve, F. L. de Hostos, E. J. Behn, G. A. Ogilvie, E. J. Powell, W. E. Pullen, W. H. Harrison, W. H. Pitkin, Jaime Sifre, Jr., Miguel Such, Frank Martinez, J. D. Dominguez, W. H. Freng.

Annual Meeting: Third Tuesday in April.

No. of Employees: Dec. 31, 1949, 769.

Offices: Tanca 261, San Juan, Puerto Rico. New York Office: 67 Broad St.

Statistics, years ended or on Dec. 31:

	1949	1948
Stations in service	31,838	32,959

Income Account, years ended Dec. 31:

	1949	1948
Total oper. rev. ..	$2,411,691	$2,250,577
Oper. expenses ..	935,817	737,603
Maintenance	463,561	401,115
Depreciation	496,591	451,925
General taxes	195,651	184,030
Income tax	96,858	125,177
Net oper. income.	240,263	267,625
Other income	8,544	dr 1,683
Total income	254,747	236,042
Interest, etc., net.	195,852	161,023
Net income	58,165	125,019
Earn. surplus, 1-1	671,234	705,443
Credits	[b]11,353	[c]93,372
Earn. surpl., 12-31	1,000,782	871,231

[a]Net adjustments applicable to prior years Puerto Rican income taxes.

[b]Credit equivalent to reduction in income tax resulting from payments during 1948 of retroactive wages and expenses applicable to final settlement of claims.

Earned per Share, common (in $), years to Dec. 31:

1949	1948	1947	1946	1945	1944	1943
3.23	0.95	5.56	0.63	1.31	2.00	5.55

No. of shares, 18,000.

Balance Sheet, as of Dec. 31:

Assets:	1949	1948
[a]Plant, prop., etc.	$5,831,931	$3,458,442
Current Assets:		
Cash	371,284	110,183
Accts. rec., net.	288,223	250,318
Mat. & supplies	261,281	254,173
Debt disc. & exp..	16,125	20,033
Defer. receivables	56,643	43,574
Prepd. & def. chg.	13,223	123,373
Total	$10,003,693	$9,267,502
Liabilities:		
Com. stock ($10)..	$1,800,000	$1,800,000
First 4½s, 1953 ...	500,000	500,000
Due to parent Co.	2,572,457	2,433,311
Curr. & Accr. Liab.:		
Long term debt	100,000	100,000
Due to affil. cos.	458,150	435,693
Accts. etc. pay.	131,403	110,566
Accr. taxes	141,516	190,271
Accr. interest ..	18,000	15,000
Empl. pens. res...	257,935	240,635
Unearn. revenues	17,503	16,811
Depreciation res..	2,623,001	2,253,919
Earned surplus ..	1,000,782	871,231
Total	$10,035,459	$9,267,502
Current assets ...	$920,796	$715,248
Curr. & accr. liab.	831,959	851,535
Net current assets	88,837	d 136,287

[a]Includes intangibles: 1949, $166,161; 1948, $165,169.

Accounts certified, Arthur Andersen & Co.

Bonded Debt L. Porto Rico Telephone Co.

First 4½s, series A; due serially to 1956:

AUTHORIZED—Series A, $1,500,000 outstanding, series A, Mar. 2, 1950, $500,000; retired, $250,000. Privately held.

MATURITY—Due $100,000 annually to Mar. 1, 1953, incl.

INTEREST—M&S 1 at company's office, 67 Broad St., New York. Principal and interest payable in U. S. legal tender.

TRUSTEE—J. P. Morgan & Co., Inc., New York. Charles R. Hatzell, San Juan, Porto Rico, Individual Trustee.

DENOMINATION—Coupon, $1,000; not registerable as to principal; fully registered, multiples of $1,000. C&R interchangeable.

CALLABLE—As a whole or in part at any time on at least 30 days' notice at par plus 1/2% for each six months or part thereof to respective serial maturities.

SINKING FUND—None.

SECURITY—A first mortgage on all properties, franchises and telephone system in Puerto Rico. Mortgage covers after-acquired properties to a limited extent.

CREATION OF ADDITIONAL DEBT—Additional bonds may be issued in other series to extent of 60% of additions to plant property after net additions of $1,000,000 have been made to plant property as of Dec. 31, 1941.

PURPOSE—Issued to repay bank loans and other loans connected with conversion of San Juan, Puerto Rico manual central offices to automatic operation and for other plant additions.

OFFERED—($1,500,000) sold privately in Mar., 1943 to Insurance companies in Canada and United States.

Capital Stock: L. Porto Rico Telephone Co. common; par $10):

AUTHORIZED—12,000 shares (increased from 12,000 shares to 20,000 shares in 1939 and to 43,000 shares in 1949); outstanding, 18,000 shares; par $100.

OWNERSHIP—As of Dec. 31, 1949, International Tel. & Tel. Corp. owned 97.69% of outstanding shares.

VOTING RIGHTS—Has one vote per share.

DIVIDENDS PAID:

1915...	$2.00	1918-19	$4.60	1920...	$8.00
1921-27	8.00	1928...	6.60	1929-36	Nil
1937-43	6.00	1944...	4.00	1945-49	Nil

TRANSFER AGENT—Stock transferred at company's office, New York.

Subscription Rights—Common stockholders of record Mar. 24, 1930, had right to subscribe for 6,000 additional common shares at $100 per share on basis of one new share for each two shares held.

Interest and Dividend Tax Ruling: The Commissioner of Internal Revenue has ruled that interest on bonds and dividends on stock of company are to be regarded, for tax purposes, as income from without the United States during the calendar year 1950. Such income, when received by non-resident aliens, is not subject to United States income tax

FIGURE B.2: A Moody's Manuals entry: The Porto Rico Telephone Company, 1949

os/annualreports/index.shtm

- Input-Output Data: Historical SIC Data, www.bea.gov/industry/io_histsic.htm
- Interstate Commerce Commission (ICC) Reports
- Moody's Manuals, 1926-1940: http://webreports.mergent.com/
- Statistics of Income: http://www.irs.gov/pub/irs-soi/
- National Association of Railroad and Utility Commissioners
- National Resources Committee (NRC) (1939), The Structure of the American Economy (Washington, DC: U.S. Government Print Office)
- Regulation of Stock Ownership in Railroads, 71st Congress, 3d Session, House Report No. 2789, Vol.2, February 1931
- Securities and Exchange Commission (SEC) Annual Reports: www.sec.gov/about/annrep.shtml
- Survey of American Listed Corporations: Reported Information on Registrants with the SEC under the Securities Exchange Act of 1934, 1939-40
- Temporary National Economic Committee (TNEC), (1940), The Distribution of Ownership in the 200 Largest Nonfinancial Corporations, monograph 29 (1-2) (Washington, DC: U.S. Government Printing Office)[3]
- Twentieth Century Fund, Committee on Taxation (1937), Facing the Tax Problem (New York: Twentieth Century Fund)

[3]http://www.bpl.org/govinfo/online-collections/federal-executive-branch/temporary-national-economic-committee-1938-1941/

B.3.2 Corporate Histories

- http://www.Archive.org
- Encyclopedia of American Business History (Facts on File, 2005): http://www.Fundinguniverse.com
- The New York Times Archives: http://www.nytimes.com/ref/membercenter/nytarchive.html
- The Wall Street Journal Archives: http://pqasb.pqarchiver.com/wsj/search.html

B.3.3 Control Chains

We use Moody's Manuals to track companies controlling, or controlled by, the 200 companies on the B&M list. In each volume, a company report is followed by reports on its controlled subsidiaries (which are identified without an explicitly specified control threshold held by the controlling company). For example, if company A controls company B and company B, in turn, controls company C, and all three firms belong to the railroad sector, the A-B-C control chain will appear in Moody's Railroads Manual in the same sequence with the identity of the corporate controller usually reported next to the company name. We examine if one or more companies are controlled by another corporation included in the original list and, if this is the case, combine their control chains. Therefore, each control chain in our sample is a long sequence of firms consisting of an apex corporation and its subsidiaries, each of which has control over the next one. In most cases, control chains include firms belonging to the same industrial category (e.g., railroads), but there are occasionally multiple control chains in different categories with the same ultimate owner as well (e.g., a few cases of public utility apex companies controlling industrial companies).

B.3.4 Ultimate Controlling Shareholders

Moody's Manuals do not provide any information on the identity of the controllers of apex firms. To identify the owners of apex corporations that are not controlled by any other entity, we use the following sources:

1. For the 1926-1929 period: Pinchot (1928), the Wall Street Journal (WSJ) and the New York Times (NYT) archives, as well as additional sources, such as internet searches, historical documents, corporate files, www.archives.org and www.fundinguniverse.com.

2. For the 1929-1932 period: Table XII, Berle and Means (1932), Bonbright and Means (1932), Buchanan (1936), Lundberg (1937), the Encyclopedia of American Business History (2006), the WSJ and NYT archives and www.fundinguniverse.com.

3. For the 1937-1940 period: National Resources Committee (1939, Chapter IX and Appendix 13) and TNEC (1940).

B.3.5 Matching Corporations to Patents, Publications and Labs

Matching Corporations to Patents

Our patent data is sourced from the Google Patents dataset via Google BigQuery.[4] We cross-check the number of utility patents granted each year with the official USPTO statistics for our sample period in Figure B.3 to ensure that our data source does not have coverage issues.[5] We find that the missing rate is around 3.43%; there are an average of 42,476 utility patents granted per year between 1926 and 1940.

[4]Please see https://cloud.google.com/blog/topics/public-datasets/google-patents-public-datasets-connecting-public-paid-and-private-patent-data for a brief overview of the dataset.

[5]USPTO official statistics for this period come from https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

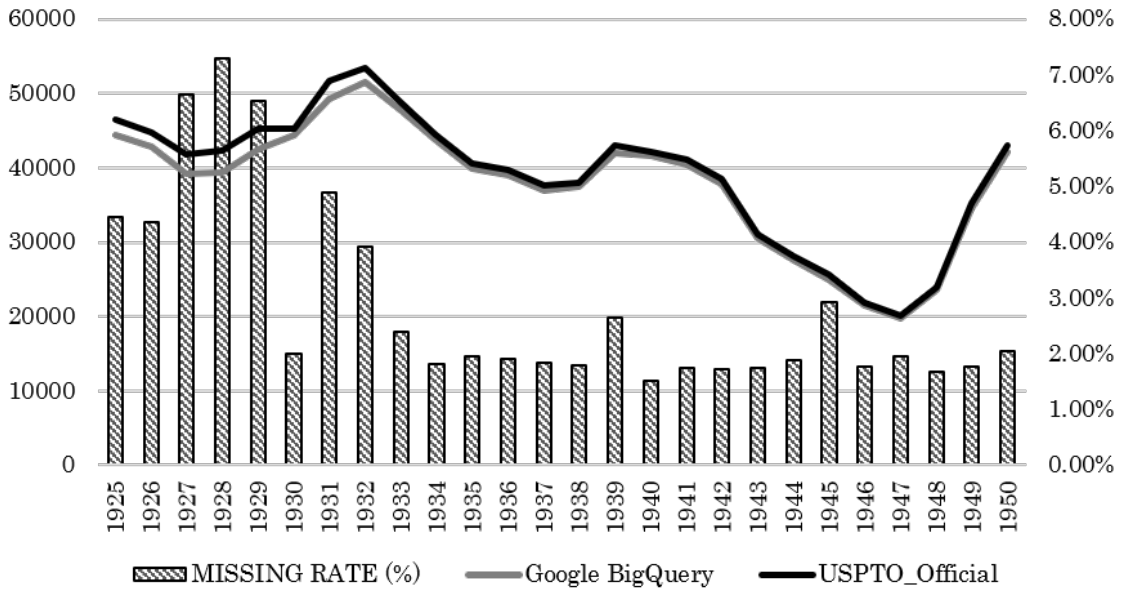


FIGURE B.3: Number of Published Utility Patents, 1925-1950[6]

We extract the assignee field of the patents and standardize the names. We remove common prefixes and suffixes, such as 'The," "LLC," "INC," "A CORP OF". We also standardize names common in certain industries such as petroleum (sometimes abbreviated as "petr"), utilities ("power" abbreviated as "pwr"), rail ("railway," "railroad," "rail" used interchangeably and variously abbreviated as "RC," "RW," "RD," and "RC") as well as more common names, such as "manufacturing" ("MFG"), "National" ("Nat'l Steel Corp."), "American" ("Radio Corp of Amer") and state abbreviations. The last standardization is important for our sample period because companies then were more often named after the states they operated in (for instance, "Delaware Lackwanna Western Coal Co." or the "Pennsylvania Electric Company"). Furthermore, we find alternative names specific certain firms such as the Standard Oil Company of Indiana (STANOLIND) and lab names for large companies such as AT&T's Bell Laboratories. Common abbreviations, such as

[6]The bar graph (right axis) plots the missing rate, defined as the difference in annual patent numbers between the USPTO official statistics and the Utility Patent (inventions) Column in the following source: https://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm.

RCA (Radio Corporation of America) and GE (General Electric), are also included. We then use a fuzzy string matching algorithm that calculates a length-adjusted Levehnstein distance. Using a fuzzy string matching algorithm is critical for patents from this period, as assignee names were not input electronically and are parsed through OCR.[7] Moreover, we manually check the names of 620 patentees with above 100 patents to include any matches that the string matching algorithm may still have missed.

We match 318 firms found in the Moody's directories to 64,523 patents. We also add 2,344 additional patents matched to 38 CRSP firms that were not matched in Kogan, Papanikolaou, Seru and Stoffman (2017).[8]

[7]As an example, the SOCONY Vacuum Oil Company is "misspelled" in the Google Patent data as: SCONY VACUUM OIL CO INC, SOCCNY VACUUM OIL CO INC, SOCENY VACUUM OIL CO IN, SOCONEY VACUUM OIL CO INC, SOCONY VACUNM OIL CO INC, SOCONY VAEUUM OIL CO INC, SOCONYVACUUM OIL CO INC, SOECNY VACUUM OIL CO INC, SOEONY VACUUM OIL CO INC, and SONCONY VACUUM OIL CO INC. The fuzzy string matching algorithm is still able to recover these matches.

[8]Kogan, Papanikolaou, Seru and Stoffman (2017) match 60,493 patents to 368 CRPS firms between 1926 and 1940, which we also add to our sample.

Table B.6: American Corporate Publications (Top 20)

Firm Name	Paper Count
General Electric Co.	919
American Telephone & Telegraph Co.	562
Westinghouse Elec. & Mfg. Co.	321
Radio Corp. of America	195
Eastman Kodak Co.	132
Humble Oil and Refining Company	45
Commonwealth Edison Co.	44
Swift & Co.	42
SHARP & DOHME INC	36
Procter & Gamble Co.	34
Western Union Tel. Co.	34
westinghouse lamp company	32
PARKE DAVIS & CO	31
Western Electric Company, Inc	30
Detroit Edison Co.	29
General Motors Corp.	28
National Carbon Co., Inc.	25
Texas Corp.	25
Aluminum Company of America	24
CORNING GLASS WORKS	24

Notes: The table presents the number of scientific publications in MAG between 1925 and 1940 matched to our sample of American firms. The top 20 publishing firms are included.

Our publication data is sourced from Microsoft Academic Graph. We first download all author affiliations for papers published between 1926 and 1940. We run the same fuzzy string matching algorithm as above and manually check matches above a threshold score. Unlike patents, corporate publications are also often published under the name of the lab, which may not always correspond to the name of the firm. Therefore, we add names of prominent corporate laboratories such as Bell Labs and the Edgar C Bain Lab (for U.S. Steel) as name variants. To prevent false positive matches, we check that charitable organizations and university labs are not

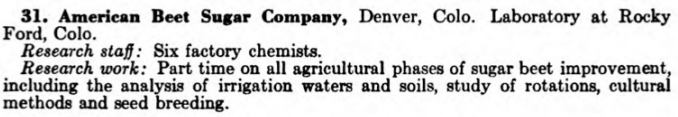
mismatched to the company. For instance, a 1934 publication by the "Eastman Laboratory of Physics" has high textual similarity to Eastman Kodak, but is actually part of the Massachusetts Institute of Technology. We also cross-tabulated the publication field of the company with its industry as a sanity check: we confirm, for instance, the wholesale and retail industry has scientific publications because the Boots Pure Drug Company (classified under this industry) published 29 articles ranging from the chemical sciences to clinical medicine.

Table B.7: American Corporate Publications, by Scientific Field

OECD Subfield	Paper Count
1.01 Mathematics	77
1.02 Computer and information sciences	77
1.03 Physical sciences and astronomy	433
1.04 Chemical sciences	267
1.05 Earth and related environmental sciences	96
1.06 Biological sciences	71
1.07 Other natural sciences	2
2.01 Civil engineering	73
2.02 Electrical eng, electronic eng	1268
2.03 Mechanical engineering	148
2.04 Chemical engineering	22
2.05 Materials engineering	152
2.06 Medical engineering	21
2.07 Environmental engineering	169
2.08 Environmental biotechnology	6
2.11 Other engineering and technologies	97
3.01 Basic medical research	23
3.02 Clinical medicine	112
3.03 Health sciences	19
4.01 Agriculture, forestry, fisheries	8
4.02 Animal and dairy science	10
4.03 Veterinary science	3
4.05 Other agricultural science	10
N/A	6386

Notes: The table tabulates the scientific fields of MAG publications from 1925 to 1940 matched to our sample of American firms.

We download the PDF files for the 1927, 1931, 1933 and 1938 editions of the NRC's Industrial Research Laboratory directory from Hathitrust. Since lab entries in the directory are of varying length (e.g., a stub for a leather company vs 2 pages for DuPont) and the fields are not sorted into metadata, the use of automated string matching algorithms is inefficient. However, since the entries are listed alphabetically, the directories are still amenable to manual matching. We enlisted two research assistants that manually searched through the directory to gather the name of the lab and the number of personnel employed at them. Though the directory also lists the type of personnel employed (e.g., chemists, physicists, etc.), these are not standardized by training or salary level, making it difficult to compare across firms. Therefore, we only use the total number of personnel as the indicator of investment in science for the analysis.



31. American Beet Sugar Company, Denver, Colo. Laboratory at Rocky Ford, Colo.
Research staff: Six factory chemists.
Research work: Part time on all agricultural phases of sugar beet improvement, including the analysis of irrigation waters and soils, study of rotations, cultural methods and seed breeding.

FIGURE B.4: 1933 IRL Entry for American Beet & Sugar Company

170. Bell Telephone Laboratories, Inc., 463 West Street, New York, N. Y. This company, a unit in the Bell Telephone System, engages in fundamental research in accordance with the research program of the American Telephone and Telegraph Company and carries out developments, designs and engineering services for the Western Electric Company, which latter company is the manufacturing unit of the Bell System.

Company officers and department heads: F. B. Jewett, President; P. Norton, Assistant to President; H. P. Charlesworth, Vice President. *Heads of functional activities:* O. E. Buckley, Director of Research; A. F. Dixon, Director of Systems Development; R. L. Jones, Director of Apparatus Development; J. G. Roberts, General Patent Attorney. *General staff:* S. P. Grace, Assistant Vice President; J. E. Moravec, Assistant Vice President; G. B. Thomas, Personnel Director; John Mills, Director of Publication.

In its functional organization the Laboratories divide into two main groups, the first of which is the technical staff including approximately 2000 research physicists, chemists, engineers, and other technicians, and the second, a somewhat smaller personnel concerned with the commercial operation of the Company and the rendering of service to the technical staff. In the second group fall such activities as the maintenance of the buildings, the operation of a well-equipped model shop, the purchase of equipment, accounting, library service, transcription, photographing, blue printing and personnel activities of education, employment and medical service.

The Laboratories carries on its technical work at the address above, and at several other locations, the most important of which are: 180 Varick Street and 480 Canal Street, New York, N. Y.; Holmdel, Deal, Summit, Whippany and Chester, N. J.

Research work: Researches in electronic physics, chemistry, magnetism, optics, radio and applied mathematics; in speech, hearing, conversion of energy between acoustic and electrical systems, the generation and modulation of electrical currents and instruments for the transmission of intelligence.

Development and design of apparatus for electrical communication, both wire and radio; studies of apparatus with a view to cost reduction either in manufacture, maintenance and repair, or through improved service; investigation of materials, maintenance of standards and methods of measurement, preparation of specifications for the manufacturer.

Development and design of communication systems combining economically for efficient operation communication apparatus and circuits, power equipment and other apparatus and circuits essential to the control, switching and supervision of communication circuits; continuing studies of current design; preparation of information necessary for manufacturer and installer.

Development and design of apparatus and investigation of materials for outside telephone plant; specification for manufacture or purchase.

Development of statistical methods of inspection and their adaptation for use by installer and manufacturer; development and application of standards of quality for communication apparatus and systems; study of inspection results; continuing study of service performance of the Laboratories' designs.

FIGURE B.5: 1933 IRL Entry for AT&T Bell Labs

B.4 Details on Scientific Gap Calculations

B.4.1 American Men of Science Directory

The AMS directory lists information on each scientist in a consistent manner: the last name is followed by the title, first name, current employment and residence and main discipline. Information on date and place of birth, alma mater, past employment and membership in professional societies follow. The final item in each entry is a detailed list of keywords that describe the focal scientist's research interests. We wish to extract i) the main discipline in which each scientist works and ii) any European degrees conferred.

The general data challenge is that the OCR on the image files, while relatively high quality, still has high error rates when classifying punctuation marks (commas,

periods and semicolons) that are essential for separating out the entries into their constituent parts. Therefore, rather than splitting the text into its constituents, we directly search for the information we need. For main disciplines, we collect 131 scientific fields from a list of deceased scientists listed at the end of the 1906 and 1921 editions of AMS.[9] We conduct regular expressions (regex) on each AMS entry to determine which disciplines correspond to each scientist.[10] We further clean this data by determining the location of the regex match: if the matched discipline occurs after the birth date (Oct. 23, 75, for Gilbert Lewis in Figure B.6), we remove the match. This prevents descriptions for research interests that occur later ("Thermodynamic theory and its application to chemistry; ... ; entropy of elements; third law of thermodynamics") from matching as the main discipline in which the scientist works. In the case of Gilbert Lewis, we prevent terms such as "electro-magnetic theory" or "non-Newtonian mechanics" to match with stemmed tokens for "Electrical Engineering" and "Mechanical Engineering." Afterwards, we manually map disciplines found in AMS into their equivalents in OECD subfields.

For alma maters and professional experience, we collect the list of all universities that were active in Europe between 1801 and 1945 from Wikipedia,[11] which in turn is heavily based on Rüegg (2004). Similar to main disciplines, we use regular expressions to determine whether each entry contains a match to at least one of these universities.[12] We further clean this data by removing matches for migrants

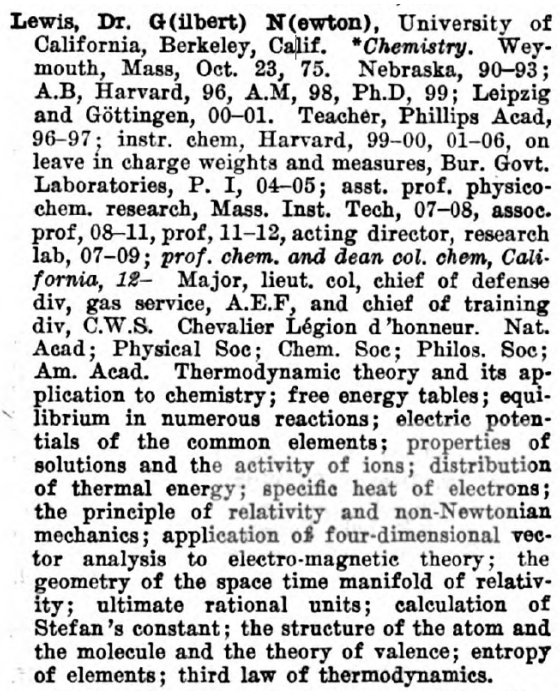
[9]These are more feasible to collect manually, as the entries are structured as names, discipline, years of birth and death.

[10]"Technology," "General Science," and "Engineering" without specifying a field (mechanical, civil, mechanical, chemical) are excluded, as they are too general.

[11]https://en.wikipedia.org/wiki/List_of_modern_universities_in_Europe_(1801\OT1\ textendash1945). We also collect data on early modern universities (established between 1501-1800) (https://en.wikipedia.org/wiki/List_of_early_modern_universities_in_Europe) and medieval universities (established before 1500) (https://en.wikipedia.org/wiki/List_of_medieval_universities) that were likely active in the early twentieth century.

[12]Schools with very short names, such as the University of Pau (France, 1722) and Literary University of Vic (Spain, 1599) are excluded because of high false positive match rates.

that were born in Europe but trained exclusively in America: any match that occurs before the birth date of the scientist is excluded. Even after this cleaning, there will remain cases where an American is trained (until his doctoral degree) in the United States, only to be recognized by foreign institutions. We therefore complement this with scientific publication output data from Clarivate Web of Science.



Lewis, Dr. G(ilbert) N(ewton), University of California, Berkeley, Calif. **Chemistry.* Weymouth, Mass, Oct. 23, 75. Nebraska, 90–93; A.B, Harvard, 96, A.M, 98, Ph.D, 99; Leipzig and Göttingen, 00–01. Teacher, Phillips Acad, 96–97; instr. chem, Harvard, 99–00, 01–06, on leave in charge weights and measures, Bur. Govt. Laboratories, P. I, 04–05; asst. prof. physico-chem. research, Mass. Inst. Tech, 07–08, assoc. prof, 08–11, prof, 11–12, acting director, research lab, 07–09; *prof. chem. and dean col. chem, California, 12–* Major, lieut. col, chief of defense div, gas service, A.E.F, and chief of training div, C.W.S. Chevalier Légion d'honneur. Nat. Acad; Physical Soc; Chem. Soc; Philos. Soc; Am. Acad. Thermodynamic theory and its application to chemistry; free energy tables; equilibrium in numerous reactions; electric potentials of the common elements; properties of solutions and the activity of ions; distribution of thermal energy; specific heat of electrons; the principle of relativity and non-Newtonian mechanics; application of four-dimensional vector analysis to electro-magnetic theory; the geometry of the space time manifold of relativity; ultimate rational units; calculation of Stefan's constant; the structure of the atom and the molecule and the theory of valence; entropy of elements; third law of thermodynamics.

FIGURE B.6: American Men of Science Entry for Gilbert Lewis (1921)[13]

B.4.2 Web of Science Affiliations Coding

For the period between 1900 to 1920, the Microsoft Academic Graph data do not record the country of publication. Also, we find that the affiliations sections rarely list the full address of the author for this period, which leads MAG to omit country data from affiliation data. We therefore rely on Clarivate Web of Science, which has previously been used for research on the impacts of World War I on scientific production (Iaria, Schwarz and Waldinger, 2018). Of 307,847 publications listed in

[13]Entry on Gilbert Lewis from the 1921 edition of the American Men of Science Directory.

Web of Science, 15% (44,356) have country data. We code each country as American, European and Rest of the World per Table B.8 and B.9. For the remaining 85% of publications without country information, we match the names of the authors to the 1906 and 1921 versions of the Cattell directory and classify those authors found in the directory as American (and the rest as European).

Table B.8: WoS Countries and Regions (1/2)

country	region	country	region	country	region	country	region
Africa	ROW	London	EUR	Switzerland	EUR	Uganda	ROW
Argentina	ROW	Malta	EUR	Syria	ROW	Rwanda	ROW
Australia	ROW	Mexico	ROW	Thailand	ROW	Ruanda Urundi	ROW
Austria	EUR	Mozambique	ROW	The Netherlands	EUR	Nigeria	ROW
Bahamas	ROW	N WALES	EUR	Turkey	ROW	Manchuria	ROW
Barbados	ROW	N Z	EUR	UK	EUR	Esthonie	EUR
Belgium	EUR	Netherlands	EUR	Ukraine	EUR	Ecudor	ROW
Belize	ROW	New Zealand	ROW	Uruguay	ROW	BURMA	ROW
Bermuda	ROW	Nicaragua	ROW	USA	USA	West Africa	ROW
Brazil	ROW	North Ireland	EUR	USSR	EUR	Ukriane	EUR
British	EUR	North Wales	EUR	Venezuela	ROW	Southern India	ROW
British East Af	EUR	Norway	EUR	W Indies	ROW	Palestine	ROW
British Hondurs	EUR	NS Wales	EUR	Wales	EUR	NY	USA
Bulgaria	EUR	Nyasaland	ROW	Western Austral	ROW	Kenya Colony	ROW
BWI	EUR	NZ	ROW	WIA	ROW	ISA	?
Canada	ROW	P I	ROW	Yemen	ROW	Iraq	ROW
CEYLON	ROW	Panama	ROW	Yugoslavia	EUR	Great Britain	EUR
Chile	ROW	Peoples R China	ROW	SUISSE	EUR	East Africa	ROW
CHINA	ROW	Persia	ROW	Finnland	EUR	Yugoslavie	EUR
Colombia	ROW	Peru	ROW	BW1	EUR	Western Samoa	ROW
Costa Rica	ROW	Philippine Isl	ROW	West Indies	ROW	Saskatchewan	ROW
Croatia	EUR	Philippine Isla	ROW	Russland	EUR	Russian Turkest	EUR
Cuba	ROW	Philippines	ROW	Prague	EUR	No Ireland	EUR
Czech Republic	EUR	Philippines Isl	ROW	Pakistan	ROW	Jugoslavia	EUR
CZECHOSLOVAKIA	EUR	Phillipine Isla	ROW	Malaysia	ROW	Johannesburg	ROW
Denmark	EUR	PI	ROW	Argentine	ROW	Inida	ROW
Egypt	ROW	Poland	EUR	Taiwan	ROW	Indien	ROW
England	EUR	Portugal	EUR	Kenya	ROW	Estonie	EUR
Federated Malay	ROW	Prussia	EUR	Bengal	ROW	Cook Islands	EUR
Fiji	ROW	Romania	EUR	Fed Malay State	ROW	BRITISH W INDIES	EUR
Finland	EUR	Russia	EUR	South America	ROW	Sri Lanka	ROW
FMS	?	S AFRICA	ROW	Philippline Isl	ROW	Siberia	EUR
France	EUR	S Australia	ROW	Morocco	ROW	Lithuania	EUR
Germany	EUR	S India	ROW	Korea	ROW	Isle Wright	EUR
Greece	EUR	S Wales	EUR	Isle Of Man	EUR	Byelarus	EUR
Guatemala	ROW	Schweden	EUR	Engalnd	EUR	British West In	EUR
Guyana	ROW	Scotland	EUR	Ecuador	ROW	Philippine	ROW
HOLLAND	EUR	Senegal	ROW	Czechoslovakio	EUR	Belgian Congo	EUR
Honduras	ROW	Siam	ROW	Czechoslovak Re	EUR	Turkestan	ROW
Hong Kong	ROW	Sierra Leone	ROW	Columbia	ROW	Tunisia	ROW
Hungary	EUR	Singapore	ROW	Trinidad	ROW	Paris	EUR
India	ROW	South Africa	ROW	Tasmania	ROW	Maroc	ROW
Ireland	EUR	South Australia	ROW	Mauritius	ROW	Hongrie	EUR
Italien	EUR	South India	ROW	Estonia	EUR	Chili	ROW
Italy	EUR	South Korea	ROW	Esthonia	EUR	Tchecoslovaquie	EUR
Jamaica	ROW	Spain	EUR	Dutch E Indies	EUR	Haiti	ROW
Japan	ROW	Sudan	ROW	Dominican Repub	ROW	Berlin	EUR
Latvia	EUR	Sweden	EUR	Union Of South	ROW	Belguim	EUR

Notes: The table lists the country affiliations of publications found in Clarivate Web of Science's Science Citation Index-Expanded between 1900 and 1920. "Region" has been imputed by the authors.

Table B.9: WoS Countries and Regions (2/2)

country	region	country	region	country	region
Union S Africa	ROW	Cihina	ROW	Paraguay	ROW
Ukraina	EUR	Britsh India	EUR	New Mexico	USA
Serbia	EUR	British India	EUR	Lebanaon	ROW
ROUMANIA	EUR	USRR	EUR	Istanbul	ROW
Polen	EUR	URSS	EUR	Isreal	ROW
Haut Congo Belg	EUR	UA	EUR	Estland	EUR
Gr Britain	EUR	Ruanda	ROW	E Indies	ROW
Chilli	EUR	R De P	ROW	E Africa	ROW
Cananda	ROW	N Ireland	EUR	Breslau	EUR
UdSSR	EUR	Dutch East Indi	EUR	Azerbaidjan	ROW
Schweiz	EUR	Czecho Solvakia	EUR	Venezuella	ROW
New Zeland	ROW	Cyprus	EUR	UL	?
Lebanon	ROW	Chechoslovakia	EUR	Slovenia	EUR
Israel	ROW	Cairo	ROW	Republic Chili	ROW
Iceland	EUR	Ukrainia	EUR	Phillipine Isl	ROW
Hawaii	USA	Scothland	EUR	Nothern Ireland	EUR
Czechoslvakia	EUR	Puerto Rico	ROW	Netherlands Ind	EUR
Abyssinia	ROW	Oslo	EUR	Lebanan	ROW
W Africa	ROW	Irlande	EUR	Georgian SSR	EUR
Sud Mandschurei	ROW	Guadeloupe	EUR	Denamrk	EUR
Porto Rico	ROW	Ethiopie	ROW	CSR	EUR
North Africa	ROW	C I	?	Britain	EUR
Netherland	EUR	BRASIL	ROW	Anglo Egyptian	EUR
Luxembourg	EUR	BELGIQUE	EUR	Yugoslavija	EUR
Irish Free Stat	EUR	Begium	EUR	Union South Afr	ROW
Iran	ROW	Bangladesh	ROW	UKx	EUR
Dominican Rep	ROW	W Germany	EUR	Sumatra	ROW
Central India	ROW	USRS	EUR	Slovakia	EUR
BWA	EUR	TH	?	RHODESIA	ROW
USAa	USA	S Africia	ROW	Northern Ireland	EUR
Ungarn	ROW	Republic Panama	ROW		
RUMANIA	EUR	Madras	ROW		
R Argentina	ROW	LURSS	EUR		
N Nigeria	ROW	Lettonia	EUR		
Mailand	ROW	Jugoslawien	EUR		
Jerusalem	ROW	Cameroon	ROW		
Czlchoslovakia	EUR	ARSSR	EUR		
Czechoslavakia	EUR	Zwitzerland	EUR		
United Kingdom	EUR	USS	USA		
Republ Libanaise	ROW	Rumanien	EUR		
Norwegen	EUR	N Rhodesia	ROW		
Northern Irelan	EUR	Czechoslovak	EUR		
Java	ROW	Cent India	ROW		
Hungry	EUR	Bolivia	ROW		
GSSR	EUR	Belgien	EUR		
Ethiopia	ROW	Armenia	EUR		
Egpyt	ROW	Uzbekistan	EUR		
Czecho Slovakia	EUR	Rep of Georgia	EUR		

Notes: The table lists the country affiliations of publications found in Clarivate Web of Science's Science Citation Index-Expanded between 1900 and 1920. "Region" has been imputed by the authors.

B.4.3 Journal Country Coding and Citation Flows

We first classify journals with non-English and non-Latin names (e.g., Zeitshcrift für Physik) as European. We also classify journals with the name "American" in it as American (e.g., the American Heart Journal). We then manually classify the remaining journals by web searches. Where a full history of the journal is available, we classify the journal's home country as the place where its publisher/publishing academic society is. For instance, "Bacteriological Reviews" is a journal that was

published by the American Society of Microbiology.[14] When publisher information is not available, we use the nationality of the founding members to classify the journal. Out of the 293 journals published between 1925 and 1940, we are able to classify 272 as American or European.

[14]https://en.wikipedia.org/wiki/Microbiology_and_Molecular_Biology_Reviews

Table B.10: Country Coding of Journals (1/7)

	Journal Name	Country
1	AMERICAN HEART JOURNAL	USA
2	AMERICAN JOURNAL OF ANATOMY	USA
3	AMERICAN JOURNAL OF BOTANY	USA
4	AMERICAN JOURNAL OF DISEASES OF CHILDREN	USA
5	AMERICAN JOURNAL OF HYGIENE	USA
6	AMERICAN JOURNAL OF HYGIENE-MONOGRAPHIC SERIES	USA
7	AMERICAN JOURNAL OF INSANITY	USA
8	AMERICAN JOURNAL OF MATHEMATICS	USA
9	AMERICAN JOURNAL OF NURSING	USA
10	AMERICAN JOURNAL OF OBSTETRICS AND GYNECOLOGY	USA
11	AMERICAN JOURNAL OF PATHOLOGY	USA
12	AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY	USA
13	AMERICAN JOURNAL OF PHYSIOLOGY	USA
14	AMERICAN JOURNAL OF PSYCHIATRY	USA
15	AMERICAN JOURNAL OF PSYCHOLOGY	USA
16	AMERICAN JOURNAL OF PUBLIC HEALTH	USA
17	AMERICAN JOURNAL OF PUBLIC HEALTH AND THE NATIONS HEALTH	USA
18	AMERICAN JOURNAL OF ROENTGENOLOGY	USA
19	AMERICAN JOURNAL OF ROENTGENOLOGY AND RADIUM THERAPY	USA
20	AMERICAN JOURNAL OF SCIENCE	USA
21	AMERICAN JOURNAL OF THE MEDICAL SCIENCES	USA
22	AMERICAN MINERALOGIST	USA
23	AMERICAN NATURALIST	USA
24	ANATOMICAL RECORD	USA
25	ANNALS OF INTERNAL MEDICINE	USA
26	ANNALS OF MATHEMATICS	USA
27	ANNALS OF SURGERY	USA
28	ARCHIVES OF INTERNAL MEDICINE	USA
29	ARCHIVES OF NEUROLOGY AND PSYCHIATRY	USA
30	ARCHIVES OF OPHTHALMOLOGY	USA
31	ARCHIVES OF OTOLARYNGOLOGY	USA
32	ARCHIVES OF PATHOLOGY	USA
33	ARCHIVES OF PATHOLOGY & LABORATORY MEDICINE	USA
34	ARCHIVES OF SURGERY	USA
35	ASTROPHYSICAL JOURNAL	USA
36	BELL SYSTEM TECHNICAL JOURNAL	USA
37	BIOLOGICAL BULLETIN	USA
38	BOSTON MEDICAL AND SURGICAL JOURNAL	USA
39	BOTANICAL GAZETTE	USA
40	BULLETIN OF THE AMERICAN MUSEUM OF NATURAL HISTORY	USA
41	BULLETIN OF THE GEOLOGICAL SOCIETY OF AMERICA	USA
42	BULLETIN OF THE JOHNS HOPKINS HOSPITAL	USA
43	BULLETIN OF THE TORREY BOTANICAL CLUB	USA
44	BUREAU OF STANDARDS JOURNAL OF RESEARCH	USA
45	CHEMICAL REVIEWS	USA
46	ECOLOGY	USA
47	ENDOCRINOLOGY	USA
48	GENETICS	USA
49	INDUSTRIAL AND ENGINEERING CHEMISTRY	USA

Table B.11: Country Coding of Journals (2/7)

	Journal Name	Country
50	JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY	USA
51	JOURNAL OF BACTERIOLOGY	USA
52	JOURNAL OF BIOLOGICAL CHEMISTRY	USA
53	JOURNAL OF BONE AND JOINT SURGERY	USA
54	JOURNAL OF CLINICAL ENDOCRINOLOGY	USA
55	JOURNAL OF CLINICAL INVESTIGATION	USA
56	JOURNAL OF COMPARATIVE NEUROLOGY	USA
57	JOURNAL OF COMPARATIVE NEUROLOGY AND PSYCHOLOGY	USA
58	JOURNAL OF COMPARATIVE PSYCHOLOGY	USA
59	JOURNAL OF ECONOMIC ENTOMOLOGY	USA
60	JOURNAL OF EDUCATIONAL PSYCHOLOGY	USA
61	JOURNAL OF EXPERIMENTAL MEDICINE	USA
62	JOURNAL OF EXPERIMENTAL PSYCHOLOGY	USA
63	JOURNAL OF FARM ECONOMICS	USA
64	JOURNAL OF GENERAL PHYSIOLOGY	USA
65	JOURNAL OF GEOLOGY	USA
66	JOURNAL OF HEREDITY	USA
67	JOURNAL OF IMMUNOLOGY	USA
68	JOURNAL OF INDUSTRIAL AND ENGINEERING CHEMISTRY-US	USA
69	JOURNAL OF INFECTIOUS DISEASES	USA
70	JOURNAL OF LABORATORY AND CLINICAL MEDICINE	USA
71	JOURNAL OF MEDICAL RESEARCH	USA
72	JOURNAL OF MORPHOLOGY	USA
73	JOURNAL OF MORPHOLOGY AND PHYSIOLOGY	USA
74	JOURNAL OF NERVOUS AND MENTAL DISEASE	USA
75	JOURNAL OF NEUROLOGY AND PSYCHOPATHOLOGY	USA
76	JOURNAL OF NUTRITION	USA
77	JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS	USA
78	JOURNAL OF PHYSICAL CHEMISTRY	USA
79	JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA	USA
80	JOURNAL OF THE AMERICAN CERAMIC SOCIETY	USA
81	JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	USA
82	JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	USA
83	JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION	USA
84	JOURNAL OF THE FRANKLIN INSTITUTE	USA
85	JOURNAL OF THE OPTICAL SOCIETY OF AMERICA	USA
86	JOURNAL OF THE OPTICAL SOCIETY OF AMERICA AND REVIEW OF SCIENTIFIC INSTRUMENTS	USA
87	JOURNAL OF UROLOGY	USA
88	NEW ENGLAND JOURNAL OF MEDICINE	USA
89	ORGANIC SYNTHESES	USA
90	PEDAGOGICAL SEMINARY	USA
91	PEDAGOGICAL SEMINARY AND JOURNAL OF GENETIC PSYCHOLOGY	USA
92	PHYSICAL REVIEW	USA
93	PHYSIOLOGICAL REVIEWS	USA
94	PHYTOPATHOLOGY	USA
95	PLANT PHYSIOLOGY	USA

Table B.12: Country Coding of Journals (3/7)

	Journal Name	Country
96	PROCEEDINGS OF THE AMERICAN ACADEMY OF ARTS AND SCIENCES	USA
97	PROCEEDINGS OF THE INSTITUTE OF RADIO ENGINEERS	USA
98	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	USA
99	PSYCHOLOGICAL BULLETIN	USA
100	PSYCHOLOGICAL REVIEW	USA
101	PUBLIC HEALTH REPORTS	USA
102	PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION	USA
103	QUARTERLY PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION	USA
104	QUARTERLY REVIEW OF BIOLOGY	USA
105	REVIEWS OF MODERN PHYSICS	USA
106	SCIENCE	USA
107	STAIN TECHNOLOGY	USA
108	TRANSACTIONS OF THE AMERICAN INSTITUTE OF CHEMICAL ENGINEERS	USA
109	TRANSACTIONS OF THE AMERICAN INSTITUTE OF MINING AND METALLURGICAL ENGINEERS	USA
110	TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY	USA
111	TRANSACTIONS-AMERICAN GEOPHYSICAL UNION	USA
112	AMERICAN JOURNAL OF MENTAL DEFICIENCY	USA
113	AMERICAN JOURNAL OF ORTHOPSYCHIATRY	USA
114	ANNALS OF MATHEMATICAL STATISTICS	USA
115	BACTERIOLOGICAL REVIEWS	USA
116	DISEASES OF THE NERVOUS SYSTEM	USA
117	ECONOMETRICA	USA
118	INDUSTRIAL AND ENGINEERING CHEMISTRY-ANALYTICAL EDITION	USA
119	JOURNAL OF APPLIED PHYSICS	USA
120	JOURNAL OF CHEMICAL PHYSICS	USA
121	JOURNAL OF CONSULTING PSYCHOLOGY	USA
122	JOURNAL OF MARINE RESEARCH	USA
123	JOURNAL OF NEUROPHYSIOLOGY	USA
124	JOURNAL OF ORGANIC CHEMISTRY	USA
125	JOURNAL OF PEDIATRICS	USA
126	JOURNAL OF RESEARCH OF THE NATIONAL BUREAU OF STANDARDS	USA
127	JOURNAL OF THE NATIONAL CANCER INSTITUTE	USA
128	PSYCHOMETRIKA	USA
129	PSYCHOSOMATIC MEDICINE	USA
130	QUARTERLY JOURNAL OF STUDIES ON ALCOHOL	USA
131	REVIEW OF SCIENTIFIC INSTRUMENTS	USA
132	SOCIOMETRY	USA
133	ANNALS OF EUGENICS	UNKNOWN
134	ARCHIVES OF DERMATOLOGY AND SYPHILOLOGY	UNKNOWN
135	CONTRIBUTIONS TO EMBRYOLOGY	UNKNOWN
136	GENETIC PSYCHOLOGY MONOGRAPHS	UNKNOWN
137	HEART-A JOURNAL FOR THE STUDY OF THE CIRCULATION	UNKNOWN
138	JOURNAL OF AGRICULTURAL RESEARCH	UNKNOWN
139	JOURNAL OF EXPERIMENTAL ZOOLOGY	UNKNOWN

Table B.13: Country Coding of Journals (4/7)

	Journal Name	Country
140	JOURNAL OF HYGIENE	UNKNOWN
141	JOURNAL OF THE INSTITUTE OF METALS	UNKNOWN
142	MEDICINE	UNKNOWN
143	MENTAL HYGIENE	UNKNOWN
144	PROCEEDINGS OF THE SOCIETY FOR EXPERIMENTAL BIOLOGY AND MEDICINE	UNKNOWN
145	PSYCHOLOGICAL MONOGRAPHS	UNKNOWN
146	SOIL SCIENCE	UNKNOWN
147	ACTA PHYSICOCHIMICA URSS	UNKNOWN
148	CHARACTER AND PERSONALITY	UNKNOWN
149	JOURNAL OF CELLULAR AND COMPARATIVE PHYSIOLOGY	UNKNOWN
150	JOURNAL OF OTOLARYNGOLOGY	UNKNOWN
151	PHYSICS-A JOURNAL OF GENERAL AND APPLIED PHYSICS	UNKNOWN
152	PSYCHIATRY	UNKNOWN
153	SURGERY	UNKNOWN
154	HELVETICA CHIMICA ACTA	SWITZERLAND
155	ACTA MEDICA SCANDINAVICA	SWEDEN
156	HEREDITAS	SWEDEN
157	ZOOLOGISKA BIDRAG FRAN UPPSALA	SWEDEN
158	ACTA MATHEMATICA	SWEDEN
159	COMPTES RENDUS DE L ACADEMIE DES SCIENCES DE L URSS	RUSSIA
160	JOURNAL OF PHYSICS-USSR	RUSSIA
161	ACTA RADIOLOGICA	NORWAY
162	PHYSICA	NETHERLANDS
163	PHYSICA B-CONDENSED MATTER	NETHERLANDS
164	PROCEEDINGS OF THE KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN	NETHERLANDS
165	PROCEEDINGS OF THE KONINKLIJKE AKADEMIE VAN WETENSCHAPPEN TE AMSTERDAM	NETHERLANDS
166	RECUEIL DES TRAVAUX CHIMIQUES DES PAYS-BAS	NETHERLANDS
167	ANNALS OF APPLIED BIOLOGY	GREAT BRITAIN
168	ANNALS OF BOTANY	GREAT BRITAIN
169	BIOCHEMICAL JOURNAL	GREAT BRITAIN
170	BIOLOGICAL REVIEWS AND BIOLOGICAL PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY	GREAT BRITAIN
171	BIOMETRIKA	GREAT BRITAIN
172	BMJ-BRITISH MEDICAL JOURNAL	GREAT BRITAIN
173	BRAIN	GREAT BRITAIN
174	BRITISH JOURNAL OF DERMATOLOGY	GREAT BRITAIN
175	BRITISH JOURNAL OF DERMATOLOGY AND SYPHILIS	GREAT BRITAIN
176	BRITISH JOURNAL OF EXPERIMENTAL BIOLOGY	GREAT BRITAIN
177	BRITISH JOURNAL OF EXPERIMENTAL PATHOLOGY	GREAT BRITAIN
178	BRITISH JOURNAL OF MEDICAL PSYCHOLOGY	GREAT BRITAIN
179	BRITISH JOURNAL OF SURGERY	GREAT BRITAIN

Table B.14: Country Coding of Journals (5/7)

	Journal Name	Country
180	BRITISH MEDICAL JOURNAL	GREAT BRITAIN
181	JOURNAL OF ANATOMY	GREAT BRITAIN
182	JOURNAL OF ANATOMY AND PHYSIOLOGY	GREAT BRITAIN
183	JOURNAL OF ECOLOGY	GREAT BRITAIN
184	JOURNAL OF GENETICS	GREAT BRITAIN
185	JOURNAL OF MENTAL SCIENCE	GREAT BRITAIN
186	JOURNAL OF PATHOLOGY AND BACTERIOLOGY	GREAT BRITAIN
187	JOURNAL OF PHYSIOLOGY-LONDON	GREAT BRITAIN
188	JOURNAL OF THE CHEMICAL SOCIETY	GREAT BRITAIN
189	JOURNAL OF THE ROYAL STATISTICAL SOCIETY	GREAT BRITAIN
190	LANCET	GREAT BRITAIN
191	MEMOIRS OF THE ROYAL METROLOGICAL SOCIETY	GREAT BRITAIN
192	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY	GREAT BRITAIN
193	NATURE	GREAT BRITAIN
194	PHILOSOPHICAL MAGAZINE	GREAT BRITAIN
195	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL OR PHYSICAL CHARACTER	GREAT BRITAIN
196	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-CONTAINING PAPERS OF A BIOLOGICAL CHARACTER	GREAT BRITAIN
197	PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY	GREAT BRITAIN
198	PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY-BIOLOGICAL SCIENCES	GREAT BRITAIN
199	PROCEEDINGS OF THE LONDON MATHEMATICAL SOCIETY	GREAT BRITAIN
200	PROCEEDINGS OF THE PHYSICAL SOCIETY	GREAT BRITAIN
201	PROCEEDINGS OF THE PHYSICAL SOCIETY OF LONDON	GREAT BRITAIN
202	PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON	GREAT BRITAIN
203	PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-CONTAINING PAPERS OF A MATHEMATICAL AND PHYSICAL CHARACTER	GREAT BRITAIN
204	PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES B-CONTAINING PAPERS OF A BIOLOGICAL CHARACTER	GREAT BRITAIN
205	PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON	GREAT BRITAIN
206	QUARTERLY JOURNAL OF EXPERIMENTAL PHYSIOLOGY	GREAT BRITAIN
207	QUARTERLY JOURNAL OF MEDICINE	GREAT BRITAIN
208	QUARTERLY JOURNAL OF MICROSCOPICAL SCIENCE	GREAT BRITAIN
209	QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY	GREAT BRITAIN
210	SURGERY GYNECOLOGY & OBSTETRICS	GREAT BRITAIN
211	TRANSACTIONS OF THE FARADAY SOCIETY	GREAT BRITAIN
212	BIOLOGICAL REVIEWS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY	GREAT BRITAIN
213	BRITISH HEART JOURNAL	GREAT BRITAIN
214	CLINICAL SCIENCE	GREAT BRITAIN
215	JOURNAL OF ANIMAL ECOLOGY	GREAT BRITAIN
216	JOURNAL OF EXPERIMENTAL BIOLOGY	GREAT BRITAIN
217	JOURNAL OF NEUROLOGY AND PSYCHIATRY	GREAT BRITAIN
218	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES	GREAT BRITAIN
219	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES A-MATHEMATICAL AND PHYSICAL SCIENCES	GREAT BRITAIN
220	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON SERIES B-BIOLOGICAL SCIENCES	GREAT BRITAIN

Table B.15: Country Coding of Journals (6/7)

	Journal Name	Country
221	PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON SERIES A-MATHEMATICAL AND PHYSICAL SCIENCES	GREAT BRITAIN
222	PROCEEDINGS OF THE ROYAL SOCIETY SERIES B-BIOLOGICAL SCIENCES	GREAT BRITAIN
223	PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON SERIES A-GENERAL AND EXPERIMENTAL	GREAT BRITAIN
224	PROCEEDINGS OF THE ZOOLOGICAL SOCIETY OF LONDON SERIES B-SYSTEMATIC AND MORPHOLOGICAL	GREAT BRITAIN
225	QUARTERLY JOURNAL OF EXPERIMENTAL PHYSIOLOGY AND COGNATE MEDICAL SCIENCES	GREAT BRITAIN
226	JOURNAL OF AGRICULTURAL SCIENCE	GREAT BRITAIN
227	BIOCHEMISCHE ZEITSCHRIFT	GERMANY
228	HOPPE-SEYLERS ZEITSCHRIFT FUR PHYSIOLOGISCHE CHEMIE	GERMANY
229	KOLLOID-ZEITSCHRIFT	GERMANY
230	MATHEMATISCHE ZEITSCHRIFT	GERMANY
231	PHYSIKALISCHE ZEITSCHRIFT	GERMANY
232	ZEITSCHRIFT DES VEREINES DEUTSCHER INGENIEURE	GERMANY
233	ZEITSCHRIFT FUR ANGEWANDTE MATHEMATIK UND MECHANIK	GERMANY
234	ZEITSCHRIFT FUR ANORGANISCHE CHEMIE	GERMANY
235	ZEITSCHRIFT FUR ANORGANISCHE UND ALLGEMEINE CHEMIE	GERMANY
236	ZEITSCHRIFT FUR BIOLOGIE	GERMANY
237	ZEITSCHRIFT FUR DIE GESAMTE NEUROLOGIE UND PSYCHIATRIE	GERMANY
238	ZEITSCHRIFT FUR ELEKTROCHEMIE	GERMANY
239	ZEITSCHRIFT FUR ELEKTROCHEMIE UND ANGEWANDTE PHYSIKALISCHE CHEMIE	GERMANY
240	ZEITSCHRIFT FUR KRISTALLOGRAPHIE	GERMANY
241	ZEITSCHRIFT FUR KRYSTALLOGRAPHIE UND MINERALOGIE	GERMANY
242	ZEITSCHRIFT FUR PHYSIK	GERMANY
243	ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE–STOCHIOMETRIE UND VERWANDTSCHAFTSLEHRE	GERMANY
244	ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-ABTEILUNG A-CHEMISCHE THERMODYNAMIK KINETIK ELEKTROCHEMIE EIGENSCHAFTSLEHRE	GERMANY
245	ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-ABTEILUNG B-CHEMIE DER ELEMENTARPROZESSE AUFBAU DER MATERIE	GERMANY
246	ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE-STOCHIOMETRIE UND VERWANDTSCHAFTSLEHRE	GERMANY
247	ZEITSCHRIFT FUR WISSENSCHAFTLICHE ZOOLOGIE	GERMANY
248	ANGEWANDTE CHEMIE	GERMANY
249	ANNALEN DER PHYSIK	GERMANY
250	ARCHIV FUR DERMATOLOGIE UND SYPHILIS	GERMANY
251	ARCHIV FUR DIE GESAMTE PHYSIOLOGIE DES MENSCHEN UND DER TIERE	GERMANY
252	ARCHIV FUR ENTWICKLUNGSMECHANIK DER ORGANISMEN	GERMANY
253	ARCHIV FUR EXPERIMENTELLE PATHOLOGIE UND PHARMAKOLOGIE	GERMANY
254	ARCHIV FUR EXPERIMENTELLE ZELLFORSCHUNG	GERMANY
255	ARCHIV FUR MIKROSKOPISCHE ANATOMIE	GERMANY
256	ARCHIV FUR MIKROSKOPISCHE ANATOMIE UND ENTWICKLUNGSGESCHICHTE	GERMANY
257	ARCHIV FUR MIKROSKOPISCHE ANATOMIE UND ENTWICKLUNGSMECHANIK	GERMANY
258	ARCHIV FUR PATHOLOGISCHE ANATOMIE UND PHYSIOLOGIE UND FUR KLINISCHE MEDICIN	GERMANY
259	ARCHIV FUR PSYCHIATRIE UND NERVENKRANKHEITEN	GERMANY
260	BEITRAGE ZUR PATHOLOGISCHEN ANATOMIE UND ZUR ALLGEMEINEN PATHOLOGIE	GERMANY

Table B.16: Country Coding of Journals (7/7)

	Journal Name	Country
261	BERICHTE DER DEUTSCHEN CHEMISCHEN GESELLSCHAFT	GERMANY
262	DERMATOLOGISCHE WOCHENSCHRIFT	GERMANY
263	DEUTSCHE MEDIZINISCHE WOCHENSCHRIFT	GERMANY
264	FORTSCHRITTE DER NEUROLOGIE UND PSYCHIATRIE UND IHRER GRENZGEBIETE	GERMANY
265	JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK	GERMANY
266	JOURNAL FUR PRAKTISCHE CHEMIE-LEIPZIG	GERMANY
267	JOURNAL FUR PSYCHOLOGIE UND NEUROLOGIE	GERMANY
268	JUSTUS LIEBIGS ANNALEN DER CHEMIE	GERMANY
269	MATHEMATISCHE ANNALEN	GERMANY
270	NATURWISSENSCHAFTEN	GERMANY
271	NAUNYN-SCHMIEDEBERGS ARCHIV FUR EXPERIMENTELLE PATHOLOGIE UND PHARMAKOLOGIE	GERMANY
272	NERVENARZT	GERMANY
273	PFLUGERS ARCHIV FUR DIE GESAMTE PHYSIOLOGIE DES MENSCHEN UND DER TIERE	GERMANY
274	PSYCHOLOGISCHE FORSCHUNG	GERMANY
275	SCHWEIZER ARCHIV FUR NEUROLOGIE UND PSYCHIATRIE	GERMANY
276	SITZUNGSBERICHTE DER KONIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN	GERMANY
277	SITZUNGSBERICHTE DER PREUSSICHEN AKADEMIE DER WISSENSCHAFTEN PHYSIKALISCH-MATHEMATISCHE KLASSE	GERMANY
278	SKANDINAVISCHES ARCHIV FUR PHYSIOLOGIE	GERMANY
279	VIRCHOWS ARCHIV FUR PATHOLOGISCHE ANATOMIE UND PHYSIOLOGIE UND FUR KLINISCHE MEDIZIN	GERMANY
280	WILHELM ROUX ARCHIV FUR ENTWICKLUNGSMECHANIK DER ORGANISMEN	GERMANY
281	ANNALES DE CHIMIE ET DE PHYSIQUE	FRANCE
282	ANNALES DE CHIMIE FRANCE	FRANCE
283	ANNALES MEDICO-PSYCHOLOGIQUES	FRANCE
284	ARCHIVES INTERNATIONALES DE PHARMACODYNAMIE ET DE THERAPIE	FRANCE
285	COMPTES RENDUS DES SEANCES DE LA SOCIETE DE BIOLOGIE ET DE SES FILIALES	FRANCE
286	COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES	FRANCE
287	ENCEPHALE-REVUE DE PSYCHIATRIE CLINIQUE BIOLOGIQUE ET THERAPEUTIQUE	FRANCE
288	JOURNAL DE PHYSIQUE ET LE RADIUM	FRANCE
289	REVUE NEUROLOGIQUE	FRANCE
290	ACTA PSYCHIATRICA ET NEUROLOGICA	EUROPE
291	CANADIAN MEDICAL ASSOCIATION JOURNAL	CANADA
292	RECUEIL DES TRAVAUX CHIMIQUES DES PAYS-BAS ET DE LA BELGIQUE	BELGIUM
293	TRANSACTIONS OF THE OPHTHALMOLOGICAL SOCIETY OF AUSTRALIA	AUSTRALIA

B.4.4 Comparison Between Gap Measures

Table B.17 compares the three measures of scientific "strength" (relative backwardness) from WoS and AMS. The "Ratio" columns for each measure present the number of European-affiliated scientists, European-authored papers and citations to European journals by American papers divided by the number of scientists never affiliated with a European institution, the number of American-authored papers and citations to American journals, respectively. Intuitively, these ratios can be thought of as the "gap" or "lag" that exists between European and American institutions (fields with relatively large values are those where the scientific gap between Europe and the U.S. is large). The three measures do not yield identical results. For instance, the AMS ratio is smaller in mean and variance compared to the ones based on Web of Science publications and citations. Given the lack of citations data in civil engineering and agriculture, forestry & fisheries journals, the citations-based gap measure cannot be calculated for these fields. Notwithstanding this, when we compare the AMS measure with the scientific publication-based measure ("Papers (WOS)") by their rankings, 10 out of the 12 fields where both measures have non-missing values differ by no more than three ranks (e.g., Mathematics is ranked 2nd in the AMS-based ratios, and is also ranked 2nd in the WoS-based ranking). The two outliers are basic medical research and agriculture, forestry & fisheries. For medicine, we suspect that there may be an over-representation of practitioners (i.e., practicing physicians). In agriculture, forestry, and fisheries, there may be a measurement error given that it encompasses a wide variety of fields. At the firm level, i.e., when the observations are weighted by the industries or scientific subfields of firms in the sample, the correlation between the two measures (r=0.527) is slightly higher than the correlation at the scientific field level (r=0.508), suggesting that fields with the highest mismatches between AMS and WoS are not very important in the patent classes used by our

sample firms (Figure B.7).

A direct ranking comparison between the citation-based measure and the other two measures is less feasible given the number of missing values. However, the fact that physics and chemistry have high (gap) scores, whereas clinical and medical sciences have relatively low gap scores, accords with the other measures. A notable outlier in this measure is Electrical engineering, which has a high score (5.80) partly due to low overall citations (34 citations in total throughout the 20-year period, compared to chemistry, which made 1,306 total citations).[15] Excluding this outlier, the correlation between the citations-based measure and the publications-based measure is positive (r=0.302) at the scientific field level.

[15]It is unclear whether this represents a measurement error. On the one hand, the only electrical engineering journal in print during this period (1900-20) is American ("Proceedings of the Institute of Radio Engineers"") and the only other electrical engineering journal indexed in the SCI before 1940 is the BELL SYSTEM TECHNICAL JOURNAL, which is American. Therefore, electrical engineering can be thought of as an area of American excellence. However, it is also possible that this field still relied on European science in the 1900-20 period, since 21 (62%) of the 34 citations were made to physics journals (1.03).

Table B.17: Number of Scientists and Papers, Europe vs America

	Papers (WOS)			Scientists (AMS)			Citations (WOS)		
OECD Subfield Equivalent	U.S.	Europe	Ratio	U.S.	Europe	Ratio	U.S.	Europe	Ratio
2.05 Materials engineering	3,806	15,456	4.06	82	41	0.50	143	44	0.31
1.01 Mathematics	5,334	19,556	3.67	525	229	0.44	134	71	0.53
2.07 Environmental engineering	-	-	-	68	27	0.40	-	-	-
3.02 Clinical medicine	43,007	81,883	1.90	178	65	0.37	6,017	2,200	0.37
1.03 Physical sciences and astronomy	12,802	42,719	3.34	605	219	0.36	197	665	3.38
3.03 Health sciences	5,121	9,373	1.83	63	22	0.35	1,042	336	0.32
3.01 Basic medical research	32,556	34,614	1.06	928	324	0.35	4,845	2,721	0.56
1.04 Chemical sciences	31,330	75,596	2.41	1,189	383	0.32	650	656	1.01
2.02 Electrical eng, electronic eng	125	143	1.14	177	54	0.31	5	29	5.80
1.05 Earth and related environmental sciences	7,996	1,189	0.15	561	167	0.30	369	128	0.35
1.06 Biological sciences	39,262	44,261	1.13	1,482	360	0.24	3,764	3,285	0.87
2.03 Mechanical engineering	-	-	-	134	32	0.24	-	-	-
2.01 Civil engineering	1,010	636	0.63	120	19	0.16	-	-	-
4.01 Agriculture, forestry, fisheries	2,112	3,594	1.70	385	60	0.16	-	-	-

Notes: This table presents the number of scientists (from AMS) and citation-weighted articles (from WoS) that have non-missing subject and affiliation fields. The "Ratio" column for the AMS sub-columns divides the number of European-affiliated American scientists by those unaffiliated with Europe. The rows are downward-sorted by this value. The "Ratio" column for WoS sub-columns divides the number of European-affiliated papers (published globally) divided by American-affiliated papers.

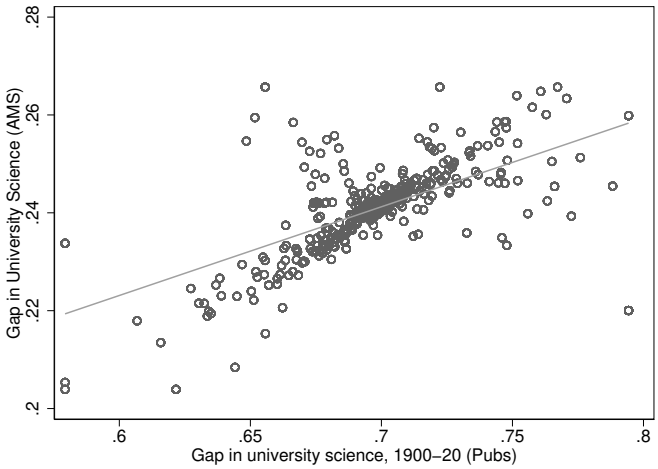


FIGURE B.7: Comparison of Gaps in University Science (1/2)[16]

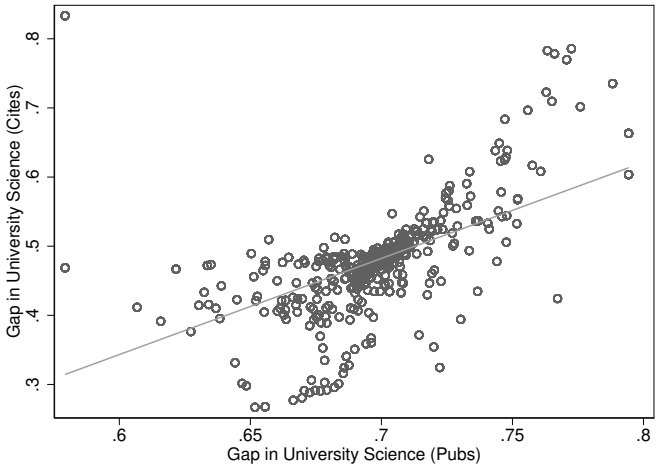


FIGURE B.8: Comparison of Gaps in University Science (2/2)[17]

[16]This figure compares the two scientific gap measures at the firm level. Higher values represent a larger gap between Europe and the United States. The AMS Scientist affiliation-based gap measure (on the vertical axis) is positively correlated with the publication volume-based gap measure (on the horizontal axis).

[17]This figure compares the two scientific gap measures at the firm level. Higher values represent a larger gap between Europe and the United States. The journal citation-based gap measure (on the vertical axis) is positively correlated with the publication volume-based gap measure (on the horizontal axis).

Appendix C

Appendix to Chapter 4

C.1 A Model of Technology Commercialization

I present a three-stage model to conceptualize the data generating process:

1. Stage 0 (Experiment): Firm experiments by paying a cost of experimentation (C). The probability of success is p. If the firm succeeds, it observes the value of the innovation x (0 otherwise).

2. Stage 1 (Entry): Given an innovation of value x, firm makes entry decision. Cost of entry is Z. Payoff from entering is $x + \epsilon$ where ϵ is a random variable. ϵ is not observed in Stage 1.

3. Stage 2 (Exit): If firm decides to enter, ϵ is realized. Firm exits if $x + \epsilon < 0$. Therefore, expected payoff given entry is $x + \mathbb{E}[\epsilon|\epsilon > -x] = X$

The market level shock ϵ follows a uniform distribution ($\epsilon \sim Uniform[-a, a]$); the value of the innovation x follows a uniform distribution with mean y and range $2b$ ($x \sim Uniform[y - b, y + b]$). I assume that $0 \leqslant x \leqslant b + y < a$ such that there is always a positive probability of exit.[1]

The firm makes two decisions:

- Firm will decide to enter at Stage 1 if $p(\mathbb{E}[X|X > Z]Pr(X > Z) - Z) - C > 0$.
- Firm will decide to experiment at Stage 0 if $p(\mathbb{E}[X] - Z) - C > 0$

C.1.1 Payoffs and Probabilities of Experimentation, Entry, and Exit

Expected payoff conditional upon entry is

$$x + \int_{-x}^{a} \frac{\epsilon}{2a} d\epsilon = x + \left[\frac{\epsilon^2}{4a}\right]_{-x}^{a} = x + \frac{a^2 - x^2}{4a} \tag{C.1}$$

[1] For $x > a$, the probability of exit would be zero since $x - \epsilon$ must be positive.

Based on these expected entry payoffs, the firm is able to estimate expected payoff from the experiment.

Expected payoff for experiment is

$$\Pi = p\left[\int_{Z}^{b+y} \frac{1}{2b}(\frac{a}{4} + x - \frac{x^2}{4a})dx - Z\right] - C^2 \tag{C.2}$$

Setting $Y = 0$ for simplicity, this reduces to

$$\Pi = p(\frac{a}{8} + \frac{b}{4} - \frac{b^2}{24a} - \frac{aZ}{8b} - \frac{Z^2}{4b} + \frac{Z^3}{24ab}) - C \tag{C.3}$$

The probability of entry is the probability that x will be above a threshold value x^*. The firm enters if $x + \mathbb{E}[\epsilon|\epsilon > -x]Pr(\epsilon + x > 0)$:

$$\begin{aligned} x + \mathbb{E}[(\epsilon|\epsilon > -x)]Pr(\epsilon > -x) &> Z \\ x + \frac{\int_{-x}^{a} \frac{\epsilon}{2a} d\epsilon}{\frac{a+x}{2a}} \frac{a+x}{2a} &> Z \\ x + \frac{a^2 - x^2}{4a} - Z &> 0 \end{aligned} \tag{C.4}$$

Setting the left term to zero, the roots are: $x^* = 2a \pm \sqrt{a^2 + 4a(a - Z)}$. Since $y + b < a$ and $x \sim Unif(y - b, y + b)$, the threshold value of x for entry is $x^* = 2a - \sqrt{a^2 + 4a(a - Z)}$,[3] which is increasing in Z: the entry threshold is higher for small firms (firms with higher Z).

[2] p is not an abbreviation for Pr but a constant for probability of experimentation success.

[3] Note that x^* is lower-bounded at $y - b$ since $x^* \in [y - b, y + b]$. For instance, for $Z = 0$, the x is zero instead of $a(2 - \sqrt{5})$

The probability of entry is given by $1 - F(x^*)$:

$$\begin{aligned} 1 - F(x^*) &= \int_{x^*}^{b+y} f(x)dx \\ &= \int_{x^*}^{b+y} \frac{1}{2b}dx \\ &= \frac{1}{2b}(b + y - 2a + \sqrt{5a^2 - 4aZ})^4 \end{aligned} \tag{C.5}$$

Probability of exit conditional upon entry is given by the following:

$$\begin{aligned} &Pr(Exit|Entry) = \mathbb{E}_x[Pr(\epsilon + x < 0|x > x^*)]^5 \\ &= \int_{x^*}^{b+y} \frac{Pr(\epsilon < -x)f(x)dx}{1 - F(x^*)} \\ &= \int_{x^*}^{b+y} \frac{a - x}{2a}\frac{1}{2b}dx \times \frac{1}{1 - F(x^*)} \\ &= \frac{1}{4ab}\int_{x^*}^{b+y}(a - x)dx \times \frac{2b}{b - x^* + y} \\ &= \frac{1}{2a(b - x^* - y)}\int_{x^*}^{b+y}(a - x)dx \\ &= \frac{5a^2 - (b + y)^2 - 2a(2Z - b - y + \sqrt{5a^2 - 4aZ})}{4a(-2a + b + y + \sqrt{5a^2 - 4aZ})} \end{aligned} \tag{C.6}$$

C.1.2 Key findings

- Payoffs from experimentation are decreasing in Z and increasing in b.
- In so far as there are tradeoffs between y and b, small firms (high Z firms) are more likely to choose novel (low y, high b) technology.
- Probability of entry and exit decrease in Z.

Derivative of experimentation payoffs with respect to Z:

$$\frac{d\Pi}{dZ} = \frac{Z^2}{8ab} - \frac{Z}{2b} - \frac{a}{8b} < 0$$
$$if \tag{C.7}$$
$$Z < a(2 + \sqrt{5})$$

Experimentation payoffs are decreasing in Z if the Z threshold condition above is satisfied. Note that Z cannot be larger than $x < y + b < a$ because that implies that the commercialization cost is larger than the highest value of the innovation. Since $Z < a$, it satisfies the threshold condition in Equation C.7.

Derivative of experimentation payoffs with respect to b: All else equal, firms prefer more novel products.

$$\frac{d\Pi}{db} = \frac{1}{b^2}\left[-\frac{Z^3}{24} + \frac{Z^2}{4} + \frac{aZ}{8} + \frac{y^3}{24a} - \frac{y^2}{4} - \frac{ay}{8}\right] + \left[\frac{1}{4} - \frac{b}{12a} - \frac{y}{8a}\right] \tag{C.8}$$

Letting $y = 0$. this reduces to:

$$= \frac{1}{b^2}\left[-\frac{Z^3}{24} + \frac{Z^2}{4} + \frac{aZ}{8}\right] + \left[\frac{1}{4} - \frac{b}{12a}\right] \tag{C.9}$$

The term without the $\frac{1}{b^2}$ multiplied is larger than zero because $b < a$. Therefore, this derivative is positive if the term inside the first bracket is positive. We know that this term is equal to zero if $Z = 0$. Moreover, it is increasing in Z for the threshold values in Equation C.7 ($Z < a(2 + \sqrt{5})$):

$$\frac{d(-\frac{Z^3}{24} + \frac{Z^2}{4} + \frac{aZ}{8})}{dZ} = -\frac{Z^2}{8} + \frac{Z}{2} + \frac{a}{8} \tag{C.10}$$

From Equation C.7 we know that $-\frac{Z^2}{8ab} + \frac{Z}{2b} + \frac{a}{8b} > 0$. The term above is a positive multiple of this value ($a, b > 0$) and therefore positive. Therefore, C.10 is positive (*all else equal*, it is better to have more variance).

Cross partial of experimentation payoffs with respect to b:

$$\frac{\partial\partial\Pi}{\partial Z \partial b} = -\frac{(Z-2a)^2 - 5a^2}{8ab^2} > 0 \tag{C.11}$$

The smallest acceptable value for Z is 0, for which the cross partial is positive. The largest possible value for Z is $a(2+\sqrt{5})$, for which the cross partial is zero. For any value in between, the cross partial is positive.

Firm preferences for b when experimenting: Assume there is a tradeoff between y and b. Let there be two firms, 0 (large) and 1 (small), with $Z_0 < Z_1$.[6] and two points for which firm 0 is indifferent between, (b, y) and (b', y') ($\Pi_0(b, y) = \Pi_0(b', y')$). The first point has higher novelty than the other point ($b > b', y < y'$). It can be shown that $\Pi_1(b, y) > \Pi_1(b', y')$:

$$\begin{aligned} \Pi_0(b, y) - \Pi_1(b, y) &< \Pi_0(b', y') - \Pi_1(b', y') \\ -\Pi_1(b, y) &< \Pi_1(b', y') \\ \Pi_1(b, y) &> \Pi_1(b', y') \end{aligned} \tag{C.12}$$

All else equal, from the same set of technologies, small firms choose the higher b option. Small firms experiment with more novel options.

[6] Z is a measure of commercialization cost, and therefore higher for smaller firms and newer markets.

Probability of Entry and Z Differentiating the result from Equation C.5 with respect to Z:

$$\begin{aligned}\frac{\partial(1-F(x^*))}{\partial Z} &= \frac{\frac{\sqrt{5a^2-4aZ}}{2b}}{\partial Z} \\ &= \frac{1}{4b}\frac{-4a}{\sqrt{5a^2-4aZ}} \\ &= -\frac{a}{b}\frac{1}{\sqrt{5a^2-4aZ}} < 0\end{aligned} \tag{C.13}$$

Therefore, the probability of entry is decreasing in Z. Probability of entry is lower for smaller firms with higher commercial uncertainty.

Probability of Exit and Z Differentiating Equation C.6 with respect to Z:

$$\frac{\partial\frac{5a^2-(b+y)^2-2a(2Z-b-y+\sqrt{5a^2-4aZ})}{4a(-2a+b+y+\sqrt{5a^2-4aZ})}}{\partial Z} = -\frac{1}{2\sqrt{5a^2-4aZ}} < 0 \tag{C.14}$$

The probability of exit decreases in Z. Higher Z firms have higher x^* which decreases their probability of entry. However, conditional upon entry, they exit less. Conversely, firms with no commercialization cost (Z) will enter at all values of x (very low x^*) and therefore will have higher probability of exit.

C.2 Historical Details on Laser Science in the Soviet Union

The Soviet Union has been home to many prominent physicists such as Lev Landau, who were supported extensively by the state. Of the USSR's ten Nobel Laureates in the natural sciences, seven were awarded to physicists. Working under the influence of Landau, Soviet physics specialized in condensed matter theory and quantum electronics (Graham, 1993). Laser physics, as a subfield of quantum electronics, was an area where Soviet scientists excelled.

Scientists in the USSR had achieved stimulated emission in the visible range independently from progress made in the West. The ideas for achieving stimulated

emission had been presented by Nikolai Basov and Alexander Prokhorov in May 1952 at the All-Soviet-Union Conference on Radiospectroscopy. The authors suggested using cesium fluoride instead of ammonia as the gain medium, which was a different solution from what the Townes & Schawlow team had been working on at Columbia and AT&T. Hence, Basov and Prokhorov received the Nobel Prize in physics in 1964 together with Townes and Schawlow for their work.

Basov had also proposed using semiconductors for lasing, once again independently from Bell Labs' Ali Javan. An "Author's Certificate" (the Soviet equivalent of a patent) pursuing this idea was granted to Basov and and Yu Popov in 1958, which preceded Robert Hall's invention of the Gallium Arsenide diode (semiconductor) laser in 1962, the first Western invention in this space (Basov and Guenther, 1985; Karlov et al., 2010). Subsequently, Prokhorov and Basov directed laser research programs around the USSR with extensive state support, and Soviet scientists were credited with the first working excimer laser in 1971 (Basov, Danilychev, Popov) as well as the synthesis of quantum dots in 1981 (Alexey Ekimov). Following up on Basov's work on semiconductor lasers, Zhores Alferov introduced the first heterostructure semiconductor lasers which opened the avenue toward room-temperature diode lasers that we use today. Therefore, Soviet laser research was overall not far behind its American counterpart. Soviet researchers achieved parity in select sectors such as semiconductor lasers and superiority in others such as rare-earth doped solid state lasers.

The cause of this divergence of focus is attributable to historical circumstances owing to the organization of laser research in the two sides. There were some similarities in that military applications were a motivator for high-energy laser research. In the United States, the Strategic Defense Initiative during the Reagan administration researched laser weapons that would intercept ballistic missiles. In the USSR, project Omega was started in 1965 also for the purposes of air defense. However,

there were important differences that caused diverging paths of competencies across the two sides of the iron curtain. In contrast to the American system that relied on a combination of government research centers such as Lawrence Livermore, large corporate R&D labs such as AT&T Bell Labs and startups such as Coherent and Spectra-Physics, the centers of laser research in the Soviet Union were concentrated in a handful of state-sponsored institutes (the Lebedev Physical Institute, Moscow state university, the Russian Academy's Institute of Spectroscopy, the Institute of Laser Physics in Siberia, the Ioffe Physico-Technical Institute, and the Belorussian Institute of Physics). Under such a centralized system, opening radically new areas of research may have been harder, but funding for existing streams of research may have been steadier. For instance, both the US and the USSR had initiated glass laser research, but the Soviet Union was the first to develop glasses capable of maintaining stable temperatures necessary for high power, pulsed lasing conditions ("athermalization"). Martin Stickley, the leading laser scientist at the Air Force, DARPA, and DOE, recalled that once the Department of Defense lost interest, no other agency was ready to champion the research:

> Interviewer (Robert Seidel): "It seems the Soviet Union in this period did not lose interest in large glass lasers for these applications, so they had developed things like athermalized glasses, which made glass lasers look more interesting not only for –"
>
> Martin Stickley: "I think the athermalized glass is something that the US could have done. I think the US can move its programs much more flexibly and quickly, and at its budget level for this, certainly chose to support gas laser technology. The Soviet Union, as you know, once something gets rolling it's quite ponderous, and it can keep on going, and the glass laser enthusiasts there, I'm sure, keep the materials development

going."[7])

Areas of (comparative) Soviet excellence in laser research thus emerged at the institute level. The Lebedev institute, where both Prokhorov and Basov remained, made seminal discoveries using rare-earth element doped glasses and crystals. The Ioffe Institute led by Alferov specialized in semiconductor lasers (Bagayev et al., 2005; Prokhorov and Shcherbakov, 1991). Indeed, a bibliography of prominent laser articles published during the 1980s shows that of the 1,160 papers that are linked to an invention of a solid state laser, 366 (32%) were authored by Soviet authors, while the Soviet share in liquid dye lasers (invented by IBM's Peter Sorokin in 1966) was 4% (7 out of 198 papers) (Weber, 2000). In some areas such as rare-earth doped solid state lasers, the Soviet Union enjoyed an absolute advantage: 85% of the papers involving Ytterbium dopants, for instance, were written by Soviet authors during this period.

In spite of these achievements, Soviet laser science could only be incompletely communicated to the West during the Cold War. First, translations of Soviet journals were often sparse and delayed, and translated versions were expensive[8] and "left many questions unanswered" (Hecht, 1999, p.93). For example, Amnon Yariv (2010 National Science Medal Laureate), an expert on phase conjugate optics recalled how Soviet articles were advanced yet hard to decipher:

> "Also, the Russian – his name is Zeldovich — who explained the thing, his article was very mathematical and using the language almost of somebody not out of the laser fraternity – was not easily understood. So one doubted the whole thing [...] So I wrote a paper [...] on the problem of complex

[7]From American Institue of Physics Oral History Interview, conducted by Robert Seidel (https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4905

[8]Up to twenty-eight times the price of their Russian language originals according to some accounts (Hollings, 2016).

> conjugation [...] And I went to read back the Zeldovich paper and I saw that, indeed, he used the word "phase conjugation" - he knew that all along."[9]

Moreover, laser inventions deemed strategic in nature were intentionally kept secret by the Soviet state. For instance, out of Alexander Prokhorov's eleven published inventions, seven were censored until the USSR's collapse (Martens, 2020).

With the collapse of the Soviet Union, state censorship of these materials were relaxed, while the migration of scientists became freer. Not only did the United States relax entry requirements for Eastern bloc scientists through the Soviet Scientist Immigration Act of 1992, but the working conditions for scientists were becoming unfavorable in Russia, with high inflation and loss of state-sponsored jobs (Mirzabekov, 1993). This led to the emigration of Soviet scientists who would further reduce the tacit information barriers that existed during the Cold War (Ganguli, 2015). Some laser scientists, such as Dmitri Basov (Nikolai Basov's son), took up jobs at academia (Columbia), while others such as Dmitri Garbuzov (Alferov's colleague), founded startups (Princeton Lightwave). Many prominent laser firms count Soviet-trained inventors in their midst during the 1990s. For instance, well over 40% of patents at Cymer[10] filed from 1989 to 2015 were invented by two Russian scientists (Alexander I. Ershov, Igor V. Fomenkov). Moreover, lasers in areas of aforementioned Soviet superiority began to be commercialized in products that were fundamentally different: the Soviet expertise in rare-earth elements, for instance, was realized when Valentin Gapontsev introduced the world's first commercial fiber laser

[9]From American Institue of Physics Oral History Interview, conducted by Lisa Bromberg https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4986

[10]Cymer was the dominant Deep Ultraviolet (DUV) excimer laser firm that enabled the introduction of 256MB DRAM chip fabrication process in 1998 as semiconductor firms switched from xenon lamps to DUV lasers for photolithography.

above 10W in 1996.[11] These lasers use glass fibers doped with rare earth elements such as Erbium and Ytterbium as the gain media, and are capable of operating at higher power and efficiency compared to conventional lasers.

In short, the development of laser science was heterogeneous across the iron curtain for reasons that are exogenous to inherent technological opportunities. The Soviet Union centralized its research around several government institutes such as the Lebedev and Ioffe Institutes which refined the lasing of rare-earth doped solid state lasers and invented the double heterostructure diode laser. Of course, American laser research often caught up in areas of Soviet excellence (and vice versa) - for instance, while excimer lasers were first invented by Basov and colleagues in 1971, the bulk of subsequent excimer research was pursued by American government (Naval Research Laboratory and Sandia Labs) and corporate (Northrop and Avco Everett Labs) researchers (Basting et al., 2002).[12] This is why throughout the paper I calculate the share of aggregate research activity using a laser handbook rather than use mere "priority" in discovery as indicators of excellence (which is the standard method in the sciences to reward credit for discoveries (Merton, 1957)).

C.3 Data Appendix

C.3.1 Details on Product and Market Data from Laser Focus

[11] Gapontsev received his PhD from the Moscow Institute of Physics and Technology and founded IPG Photonics in 1990.

[12] The Soviet share of excimer lasers during the 1980s is therefore below 10 percent.

Table C.1: Laser Product Types (Construction and Beam Type by Gain Medium) (1/2)

Construction	Generic SS		Tunable SS		DPSS		Semiconductor		Fiber		Dye		Gas	
Beam Type (continuous or pulsed)	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed
Al	5													
Ar													6964	127
Ar/Kr													552	
ArF														546
Au													2	24
Ba														3
CH3F													23	24
CH3OH													32	27
CO													122	32
CO2													3639	2529
Ce					1									
Co		9												
Cr	54	1000	13	41	74	7								
Cu													6	241
Er	14	174			32	39			6	20				
Er / Cr		4												
F														166
FH													73	90
Far_IR													196	47
GaAlAs													2	
GaAs	46	6												
H														8
H2O													6	5
HCN													11	6
HF													101	74
HeAg													1	1
HeAu													1	1
HeCd													1582	
HeNe													6669	
HeXe													8	8

Notes: This tabulates number of laser products sold *Laser Focus* Buyer's Guides by construction, beam type (constituting the columns) and gain medium (constituting the rows). Tunable SS refers to tunable solid state lasers, DPSS refers to Diode Pumped Solid State lasers, Generic SS refer to all other solid state lasers. CW refers to Continuous Wave.

Table C.2: Laser Product Types (Construction and Beam Type by Gain Medium) (2/2)

Construction	Generic SS		Tunable SS		DPSS		Semiconductor		Fiber		Dye		Gas	
Beam Type (continuous or pulsed)	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed	CW	Pulsed
Ho	10	109			6	5								
I														4
Kr													1224	75
KrCl														130
KrF														665
Mn														4
N2														513
N2O													28	19
NH3													6	
Nd	4737	13485	5	45	2250	886			1					
Nd/Cr		13												
NeCu													1	1
OTHER													117	76
Pb														7
Ti	310	376	724	3425	2	25			2					
Tm	6	39			3	1			1					
Xe														18
XeBr2														11
XeCl2 / Xe2Cl		16												717
XeF2		22												422
XeHe													8	2
Yb	5	2			16	8			31	55				
excimer														2012
metalvapor														102
N/A	179	103			940	578	11137	1999	7	26	486	1927	210	1

Notes: This tabulates number of laser products sold *Laser Focus* Buyer's Guides by construction, beam type (constituting the columns) and gain medium (constituting the rows). Tunable SS refers to tunable solid state lasers, DPSS refers to Diode Pumped Solid State lasers, Generic SS refer to all other solid state lasers. CW refers to Continuous Wave.

C.3.2 Federal Laser Contracts

I source data from Belenzon and Cioaca (2021), who collect federal procurement contracts by manually downloading government contracts by year and agency from https://www.usaspending.gov (maintained by the Office of Management and Budget) and https://beta.SAM.gov (maintained by the General Services Administration). The former dataset includes all federal procurements since 2000, while the latter contains procurement data as early as the 1970s. The combined dataset spans years between 1980 and 2019 and includes the signing date, action obligation[13] in current dollars, vendor names, contracting agency and the relevant 4-digit Product and Service Codes (PSC) for federal agencies.[14] Dollar amounts are adjusted to constant 2012 dollars using GDP deflators from Louis Johnston and Samuel H. Williamson, "What Was the U.S. GDP Then?" MeasuringWorth, 2020.

I restrict the data to product procurement contracts for lasers (PSC: 5860)[15] Of the four largest government departments for which Belenzon and Cioaca (2021) collect data on (DoD, HHS, DoE, VA), the Department of Defense accounts for 99.6% of laser contracts between 1980 and 1990 and 99.8% between 1980 and 2019. The average annual dollar amount of DoD laser contracts are $33,610,232.

[13]Action obligations are "intentions" backed by a contractual agreement. The government does not release actual dollar amounts spent on contracts. After initial contract signing, the actual expenditure can increase, decrease, or stay the same. Agencies enter a "contract action" into the database whenever they know that what was initially obligated has changed. These corrections may sometimes (5%) lead to negative obligations.

[14]There are a total of 740 PSC codes for R&D procurement, 694 for product procurement, and 2969 for services procurement.

[15]Stimulated Coherent Radiation Devices, Components, and Accessories. This class includes, and is restricted to, devices the operation of which is based on principles involving the stimulated emission of radiation, and to associated components and accessories directly related to stimulated coherent radiation techniques. The methods for producing the radiation are those obtained by LASER (Light Amplification by Stimulated Emission of Radiation) and by MASER (Microwave and/or Molecular Amplification by Stimulated Emission of Radiation) techniques. Only those items specifically designed for stimulated coherent radiation applications, and having no other application, are to be included in this class. Assemblies or major systems containing such devices are excluded from this class (from March 2020 PSC Manual by the U.S. General Services Administration Federal Acquisition Services)

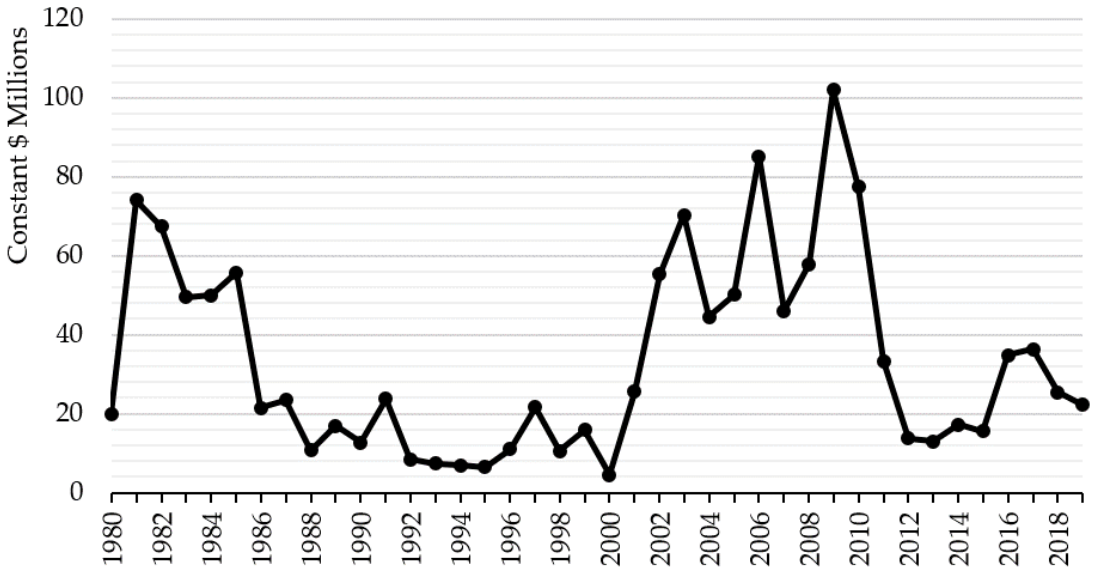


FIGURE C.1: Department of Defense Procurement Contract Obligations in Laser Products (PSC: 5860)[16]

Government Involvement and Commercialization of Laser Inventions

I present qualitative evidence documenting the role of government funding on commercialization of different types of lasers. Upon the invention of the first laser in the early 1960s, there were three main applications of lasers identified: measurement (interferometry, range finding), communication (battlefield optical links, space communication, terrestrial communication), materials processing (drilling, welding, cutting). Varying U.S. government involvement in each influenced the speed with which each application was commercialized.[17]

Measurement was an easy technical problem and had plenty of government support through the National Bureau of Standards (NBS) and the armed forces. This allowed for the fastest commercialization: the first laser was invented in 1960 – the first functioning (Helium Neon) laser distance measurement device was invented in

[16]This figure plots the total annual contract obligations of the U.S. Department of Defense for laser products between 1980 and 2019. Please see Appendix C.3.2 for details on the collection of federal laser contracts.

[17]This is based chiefly on Bromberg (1991), chapter 5

1962. The NBS continued to improve upon this and the technology was mature by 1968. The Vietnam war also produced Project 1559, a "quick reaction" fund to materialize concepts into weapons within 6-12 months. Laser-guidance missiles were developed through these programs.

Optical communication was a more difficult technical problem because laser beams either needed to travel through the atmosphere (through adverse weather) or through waveguides (which was the only technical solution until the invention of low-loss optical fiber by Corning glass in 1970). However, communication also received substantial funding from the Armed Forces and NASA, which was trying to replace microwave communications with optical ones. Gallium Arsenide Laser links commissioned by the DoD were fielded as early as 1963 (manufactured by GE, Hughes, RCA, and Sylvania GTE). Also, the potential payoff from supplanting existing telephone networks was large, which motivated heavy corporate investment into inventing (MIT, GE) and improving (AT&T) heterojunction semiconductor lasers. The first fiber optics communications were demonstrated in 1976 by Bell Labs in Atlanta.

Materials processing was also a difficult technical problem since, unlike measurement, the beam needed to interact with other substances. Lasers pulsed too slow for immediate applications in welding. Unlike communications, government interest was minimal, and progress was therefore relatively slower (practical applications lagging by about 5-10 years, according to Bromberg). Basic research on Argon Ion, Carbon Dioxide and Nd:YAG lasers needed to continue until the early 1970s, with limited applications in microprocessor manufacture (using Nd:YAG lasers, from 1967) and metal cutting for the aerospace industry (using CO2 lasers, from 1970).

C.3.3 Soviet Scientist Immigration to the United States

I collect information on the immigration of scientific and engineering personnel from the Soviet Union from the Immigration and Naturalization Services' "Immigrants Admitted to the United States" series, which contains anonymized information of all aliens who became legal permanent residents in the United States between fiscal years 1972 to 2000.[18] The dataset contains information on the country of origin, occupation, and intended area of residence within the United States at the zipcode level. I restrict myself to years between 1983 and 1997 as the Immigration & Naturalization Service followed a unified occupation coding system consisting of 29 standardized occupations (shown in figure C.2 from the Documentation for the 1996 version of the data).[19]

[18]Data is accessed through the Inter-university Consortium for Political and Social Research (ICPSR) at https://www.icpsr.umich.edu/web/ICPSR/studies/37688

[19]Missing rate for occupations is .2% (22,680 out of 10,404,482 migrants) for 1983-1997

Table C.3: Top and Bottom 5 MSAs, by Increase in Soviet S&E Migrants

	Migrant Country of Origin	Soviet Union		China & India		Rest of World	
Rank	Migration Year	Pre1990	Post1990	Pre1990	Post1990	Pre1990	Post1990
1	Los Angeles–Long Beach, CA PMSA	437	850	1459	1929	10051	8055
2	New York, NY PMSA	538	837	1927	2207	8864	13521
3	Not in MSA CA	162	432	1328	2731	6722	7946
4	San Jose, CA PMSA	68	241	1424	3134	4573	5469
5	Washington, DC–MD–VA MSA	73	227	736	1429	2812	4578
6	Chicago, IL PMSA	268	417	893	1227	2824	7016
...	...	...	...	...	...	...	...
378	Waterbury, CT MSA	2	1	1	0	18	16
379	Worcester, MA MSA	3	2	53	117	164	288
380	Youngstown–Warren, OH MSA	2	1	11	21	60	44
381	Binghamton, NY MSA	2	0	12	27	30	34
382	Not in MSA HI	2	0	44	71	324	298
383	Reading, PA MSA	3	1	17	17	43	52

Notes: Unit of analysis is at the MSA level. Columns for the Soviet Union count the number of Science and Engineering migrants in the focal MSA whose country of origin is the Soviet Union or Russia. The China & India Columns count migrants whose country of origins is Mainland China, Taiwan, or India. Rest of the World Columns refer to all other migrants. Pre and Post 1990 Columns split these numbers by whether the migrant has obtained permanent residency before (or on) 1990 and after 1990 respectively. Each MSA is ranked by the difference in migrant numbers between the post and pre periods for Soviet S&E migrants.

Occupation

For aliens qualifying for immigration based on their job skill (employment-based preference principals), occupation decribes the employment they will be performing in the United States. For all other aliens, occupation refers to the employment held in their country of last residence on in the United States. This field also contains categories for aliens not in the labor force or unemployed such as homemakers, students, and unemployed or retired.

Symbol	Occupation
ARC	Architects
ART	Writers, artists, entertainers, and athletes
ASP	Administrative support occupations, including clerical
COU	Counselors, educational and vocational
DOC	Physicians
ENG	Engineers, surveyors and mapping scientists
EXC	Executive, administrative, and managerial
FFF	Farming, forestry, and fishing occupations
HLD	Health diagnosing occupations
HLT	Health assessment and Treating occupations
HOU	Homemakers
LAB	Operators, fabricators, and laborers
LAW	Lawyers and judges
LIB	Librarians, archivists, and curators
MCS	Mathematical and computer scientists
NOT	Occupation not reported
NSC	Natural scientists
NUR	Registered nurses
PCR	Precision production, craft, and repair occupations
SER	Service occupations
SLS	Sales occupations
SSC	Social scientists and urban planners
STC	Students and/or children under age 16
SWK	Social, recreation, and religious workers
TCO	Teachers, except postsecondary
TCU	Teachers, postsecondary
TNH	Health technologists and technicians
TNO	Technologists and technicians, except health
UNR	Unemployed or retired

FIGURE C.2: INS Occupation Coding for "Immigrants Admitted to the United States" (1983-1997)[20]

I select on the following occupations codes which are related to science and engineering (they are about 2.9% of total migrants):

1. Mathematical and Computer Scientist (MCS)

2. Natural Scientists (NSC)

3. Technologist and technicians, except health (TNO)

[20]This lists the occupations recorded by the INS in p.69 of the "Documentation Package for the 1996 Immigrant Public Use Tape".

4. Engineers, surveyors and mapping scientists (ENG)

To identify immigrants from the Soviet Union, I filter out migrants whose "Country of Chargeability" (country of origin) is the "Soviet Union" or "Russia". I then link migrant destination zipcodes to 383 Metropolitan Statistical Areas using cartographic shapefiles provided by the 1990 Decennial Census. I download the historical delineation files for the 1990 Decennial Census from the Census TIGER database[21] as well as the delineation files for U.S. cities and zipcodes from the ESRI USA data available from Baruch's Geoportal website.[22] I then use the "spatial join" feature in ArcGIS Pro in order to determine which cities and zipcodes lie within focal MSAs and PMSAs. Cities and zipcodes that are not classified within an MSA or PMSA are classified by their state – e.g., "Not in MSA/PMSA California." Given this bridge between zipcodes and MSAs, immigrant destinations can be classified into MSAs. Given the bridge between city-state combinations and MSAs, firms in *Laser Focus* Buyer's Guides are also classified into MSAs. In the case of a laser firm with multiple locations, the MSA with the highest number of Soviet S&E immigration is assigned to the firm. This is based on the assumption that the firm is able to integrate information from all of its locations. This links the cities and states listed for laser firms in *Laser Focus* Buyer Guides to 70 MSAs. 342 (59%) out of the 582 firms in our sample are matched to an MSA.[23] Table C.3 lists the top and bottom five MSAs in terms of the increase in Soviet S&E personnel immigration.

[21]https://www.census.gov/geo/maps-data/data/tiger-line.html

[22]https://www.baruch.cuny.edu/confluence/display/geoportal/ESRI+USA+Data

[23]Firms with non-US headquarters such as Melles Griot, Lambda Physik and Quantel account for the missing matches

C.4 Robustness Checks

C.4.1 Auxiliary Patent-Level Results

Table C.4: Novelty of Laser Patents (CPC: H01S) and Gain Media, by Assignee Type

	DV: Patent Novelty			
	(1) All	(2) Univ/Gov	(3) Startup	(4) Compustat
Year > 1990 × Soviet Paper Share (Patent)	-0.020	1.134	-0.005	-0.022
	(0.014)	(0.379)	(0.057)	(0.020)
Soviet Paper Share (Patent)	0.019	-1.155	-0.029	0.001
	(0.013)	(0.376)	(0.053)	(0.018)
Patent Cites to Science	-0.000	-0.001	0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Forward Patent Cites	-0.001	0.023	0.002	-0.000
	(0.000)	(0.009)	(0.003)	(0.001)
Triadic Patent Dummy	0.007	-0.032	0.003	0.008
	(0.002)	(0.029)	(0.010)	(0.003)
Number of Claims	0.001	0.001	0.000	0.001
	(0.000)	(0.001)	(0.000)	(0.000)
Length of First Claim	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.770	0.770	0.792	0.790
Years	31	8	27	31
R^2	0.114	0.111	0.144	0.089
Number of Observations	8,869	160	511	3,506

Notes: Unit of analysis is the patent. *Patent Novelty* is calculated as one minus the normalized textual similarity of a patent to all U.S. prior art published from 1980 (please see (Arora et al., 2018a) for details). Sample is limited to patents in CPC: H01S between 1980 and 2010. Sample for Column 2 consists of patents by academic assignees (including university hospitals) and government assignees (including government agencies, laboratories, and branches of the armed forces). Sample for Column 3 consists of patents assigned to small entities with under 500 employees classified by section 41 of the U.S. patent act. Sample for Column 4 consists of patents assigned to Compustat firms matched by the DISCERN project. Year fixed effects are included. Standard errors are robust to arbitrary heteroscedasticity.

Table C.5: Novelty of Laser Patents (CPC: H01S), by Citations to Soviet Science

	DV: Patent Novelty			
	(1)	(2)	(3)	(4)
Cites to Soviet Science (D=0)	0.024	0.016	0.015	0.015
	(0.010)	(0.010)	(0.010)	(0.010)
Cites to Soviet Science (D=1)		0.021	0.014	0.015
		(0.004)	(0.004)	(0.004)
Cites to Soviet Science (D=2)			0.009	0.016
			(0.002)	(0.003)
Cites to Soviet Science (D=3)				-0.008
				(0.002)
Patent Cites to Science	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.762	0.762	0.762	0.762
Years	34	34	34	34
R^2	0.076	0.078	0.079	0.080
Number of Observations	15,438	15,438	15,438	15,438

Notes: Unit of analysis is the patent. *Patent Novelty* is calculated as one minus the normalized textual similarity of a patent to all U.S. prior art published from 1980 (please see (Arora et al., 2018a) for details). Sample is limited to patents in CPC: H01S between 1980 and 2010. *Citations to Soviet Science (D=0)* is equal to one if the focal patent makes a citation to a Soviet laser article, and zero otherwise. *Citations to Soviet Science (D=1)* is equal to one if the focal patent makes a citation to any scientific article that cites a Soviet laser article. *Patent Cites to Science* counts the number of citations made to any scientific article found in MAG. Standard errors are robust to arbitrary heteroscedasticity.

C.4.2 Replication of Results without SDI-related Lasers

Commercial lasers relevant to the Strategic Defense Initiative are Hydrogen Fluoride, Deuterium Fluoride, and X-ray lasers. While X-ray lasers are not identifiable through their gain media, they can be identified by their wavelengths (0-10 nanometers). At the product level, there are 169 Hydrogen Fluoride and 164 Deuterium Fluoride lasers. 1,258 lasers lase in the X-ray range. I exclude these lasers and re-run the

specifications in Tables 4.6 and 4.10.

Similarly, I exclude 785 patents containing Hydrogen Fluoride or Deteurium Fluoride in their patent documents and rerun specifications in Tables C.5 below.[24]

[24]X-ray laser patents are not excluded yet.

Table C.6: Novelty of Laser Products (excluding SDI lasers)

	DV: $\|Z\ Score\|$				
	(1) Wavelength	(2) Power/Energy	(3) Diameter	(4) Pulse Length	(5) Rep Rate
Soviet Paper Share × Year > 1990	2.639	0.702	0.653	1.413	3.134
	(0.313)	(0.397)	(0.104)	(0.747)	(0.715)
Soviet Paper Share	2.868	16.272	5.582	1.213	-1.115
	(1.110)	(1.159)	(0.263)	(0.941)	(1.070)
Average of DV	1.136	1.734	0.821	1.679	1.843
Year FEs	Yes	Yes	Yes	Yes	Yes
Laser Type FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.007	0.007	0.029	0.003	0.003
Number of Observations	64,466	65,347	40,852	17,161	21,064

Notes: Unit of analysis is the product-year. Sample is limited to products unrelated to the SDI program. Variable definitions are identical to those in Table 4.6. Standard errors are robust to arbitrary heteroscedasticity.

Table C.7: Startup Entry into Novel Products (excluding SDI lasers)

	DV: First Product = 1			
	(1) All	(2) Incumbent	(3) Startup	(4) Startup(¡50)
Year > 1990 × Soviet Paper Share	0.343	0.328	0.933	1.003
	(0.046)	(0.049)	(0.112)	(0.116)
Soviet Paper Share	0.018	0.029	-0.006	0.034
	(0.017)	(0.022)	(0.009)	(0.014)
Employees	0.000	0.000	0.001	0.003
	(0.000)	(0.000)	(0.000)	(0.001)
Average of DV	0.076	0.080	0.062	0.053
Year FEs	Yes	Yes	Yes	Yes
R^2	0.331	0.405	0.347	0.329
Number of Observations	4,783	3,691	1,092	994

Notes: Unit of analysis is the product-year. Sample is limited to products unrelated to the SDI program. Variable definitions are identical to those in Table 4.10. Standard errors are robust to arbitrary heteroscedasticity.

Table C.8: Novelty of Laser Patents (CPC: H01S), by Citations to Soviet Science (excluding SDI lasers)

	DV: Patent Novelty			
	(1)	(2)	(3)	(4)
Cites to Soviet Science (D=0)	0.024	0.016	0.015	0.015
	(0.010)	(0.010)	(0.010)	(0.010)
Cites to Soviet Science (D=1)		0.020	0.014	0.014
		(0.004)	(0.004)	(0.004)
Cites to Soviet Science (D=2)			0.009	0.015
			(0.002)	(0.003)
Cites to Soviet Science (D=3)				-0.008
				(0.002)
Patent Cites to Science	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Average of DV	0.763	0.763	0.763	0.763
Years	34	34	34	34
R^2	0.076	0.078	0.079	0.080
Number of Observations	14,907	14,907	14,907	14,907

Notes: Unit of analysis is the patent. Sample is limited to patents in CPC: H01S between 1980 and 2010. *Patent Novelty* is calculated as one minus the normalized textual similarity of a patent to all U.S. prior art published from 1980 (please see (Arora et al., 2018a) for details). *Citations to Soviet Science (D=0)* is equal to one if the focal patent makes a citation to a Soviet laser article, and zero otherwise. *Citations to Soviet Science (D=1)* is equal to one if the focal patent makes a citation to any scientific article that cites a Soviet laser article. *Patent Cites to Science* counts the number of citations made to any scientific article found in MAG. Standard errors are robust to arbitrary heteroscedasticity.

C.4.3 Interaction with Soviet S&E Immigration to American Cities

To establish whether reliance on Soviet laser science was positively associated with immigration of Soviet scientists, I regress at the MSA level the number of Soviet lasers sold by firms in the focal MSA against the cumulative number of Soviet Science & Engineering migrants entering. Columns 1 and 2 in Table C.9 show that MSAs with higher number of Soviet migrants are associated with a higher number of Soviet-type

lasers being produced and sold. Controlling for the number of migrants from other countries in Columns 3 and 4 renders the estimates on Soviet Migrants insignificant and suggests that S&E migration not only brings knowledge but also skilled human capital that may be utilized for new product development. In Table C.10 Column 1, I find that patent citations to Soviet laser journals is positively associated with the number of Soviet S&E migrants in the patenting firm's MSA. This effect dissipates, however, as the citation "distance" increases.

Table C.9: Soviet Scientist Immigration and Lasers (City Level)

	DV: No. of Soviet Lasers			
	(1)	(2)	(3)	(4)
Soviet Migrants	1.331	2.998	-7.552	-2.487
	(0.439)	(1.656)	(2.707)	(5.525)
No. of Total Lasers	0.320	0.300	0.236	0.250
	(0.020)	(0.068)	(0.027)	(0.077)
China&India Migrants			3.656	3.670
			(0.862)	(2.078)
ROW Migrants			-0.124	-0.401
			(0.246)	(0.570)
Average of DV	21.381	21.599	21.381	21.599
Year FEs	Yes	Yes	Yes	Yes
MSA FEs	No	Yes	No	Yes
R^2	0.754	0.878	0.796	0.888
Number of Observations	688	681	688	681

Notes: Unit of analysis is the MSA-year. Sample is limited to MSA-years with at least one operating laser firm from Laser Focus Buyer's Guides. *Soviet Migrants* refers to the total number of Soviet Science & Engineering migrants (in hundreds) identified in section C.3.3 since 1983 up to the focal year. *China & India Migrants* refer to migrants with country of origin in Mainland China, Taiwan, and India (in hundreds). ROW migrants refers to migrants from all other countries (in hundreds). Dependent variable is the number of lasers whose gain media are in the fourth quartile of Soviet paper shares sold by firms classified in the focal MSA. Standard errors for Columns 2 and 4 are clustered at the MSA level, while those for Columns 1 and 3 are robust to arbitrary heteroscedasticity.

Table C.10: Patent Citations to Soviet Laser Science

	DV: Citation to Soviet Laser Science			
	(1) D=0	(2) D=1	(3) D=2	(4) D=3
Soviet Migrants	0.019	-0.101	-0.149	-0.089
	(0.004)	(0.051)	(0.046)	(0.058)
China&India Migrants	0.001	0.033	0.064	0.066
	(0.001)	(0.015)	(0.008)	(0.011)
ROW Migrants	-0.002	0.001	-0.002	-0.005
	(0.000)	(0.003)	(0.004)	(0.004)
Patent Cites to Science	0.000	0.005	0.012	0.020
	(0.000)	(0.002)	(0.004)	(0.007)
Average of DV	0.001	0.007	0.022	0.041
Year FEs	Yes	Yes	Yes	Yes
4 digit IPC FEs	Yes	Yes	Yes	Yes
MSA FEs	Yes	Yes	Yes	Yes
R^2	0.015	0.123	0.239	0.298
Number of Observations	10,127	10,127	10,127	10,127

Notes: Unit of analysis is the patent. Sample is limited to patents matched to firms active in *Laser Focus*. *Soviet Migrants* refers to the total number of Soviet Science & Engineering migrants (in hundreds) identified in section C.3.3 that are matched to the patenting firm's location. Dependent variable for Column 1 (D=0) is equal to one if a patent cites an article in the Soviet laser journals listed in section 4.4.1. Dependent variable for Column 1 (D=0) is equal to one if a patent cites an article that cites an article in the Soviet laser journals. Standard errors are clustered at the MSA level.

I also test whether the results comparing startups to incumbents are stronger for cities with higher Soviet Science and Engineering migrants. Table C.11 presents results from interacting the difference-in-difference estimator with the number of Soviet S&E migrants in the focal firm's MSA. I find in Columns 1 and 4 that the higher novelty associated with Soviet-type lasers after the end of the Cold war is stronger for MSAs with more migrants. Specifically, 100 more Soviet migrants in the focal firm's MSA is associated with around a 9 times larger effect compared to the result in Table 4.6.

Table C.12 tests whether the results in Table 4.10 are stronger for lasers produced by firms in cities with more Soviet migrant populations. Column 1 shows that

the products are more likely to be first in class especially if they are produced by firms with more Soviet migrants. Comparing the coefficient on the triple interaction between Column 2 and 3 shows that this result is mostly driven by startups.

Table C.11: Novelty of Laser Products, by Soviet S&E Immigration

	DV: $\lvert Z\ Score \rvert$				
	(1) Wavelen	(2) Power/Energy	(3) Diam	(4) Pulse Len	(5) Rep Rate
Soviet Paper Share × Year >1990 × Soviet Migrants	23.842	0.027	-0.335	11.050	-0.350
	(7.799)	(0.739)	(0.356)	(3.981)	(4.784)
Soviet Paper Share × Year >1990	-0.576	0.589	0.143	-3.485	10.598
	(1.463)	(1.293)	(0.849)	(4.092)	(4.817)
Soviet Paper Share × Soviet Migrants	-24.220	-0.456	0.722	-9.039	-2.268
	(7.798)	(0.772)	(0.376)	(1.777)	(5.540)
Year > 1990 × Soviet Migrants	-7.225	0.327	0.097	-1.678	-1.042
	(1.432)	(0.210)	(0.216)	(0.525)	(1.189)
Soviet Paper Share	10.818	9.028	1.882	5.247	-8.696
	(3.514)	(2.741)	(0.622)	(2.696)	(4.285)
Soviet Migrants	7.626	-0.516	-0.259	3.048	2.703
	(1.509)	(0.297)	(0.268)	(0.860)	(2.237)
China&India Migrants	0.050	-0.043	-0.001	-0.145	0.061
	(0.024)	(0.019)	(0.009)	(0.071)	(0.093)
ROW Migrants	-0.028	0.027	0.003	-0.002	-0.133
	(0.017)	(0.012)	(0.006)	(0.045)	(0.101)
Average of DV	1.142	1.508	0.720	1.473	1.949
Year FEs	Yes	Yes	Yes	Yes	Yes
Construction-Beam Type FEs	Yes	Yes	Yes	Yes	Yes
MSA FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.065	0.099	0.093	0.103	0.040
Number of Observations	42,822	43,782	28,647	9,562	13,195

Notes: Unit of analysis is the product-year. $Year$>1990 is equal to one if the panel year is larger than 1990. *Soviet Migrants* refers to the cumulative number of S&E migrants from the Soviet Union that arrive at the focal firm's MSA-year (in hundreds). Standard errors are robust to arbitrary heteroscedasticity.

Table C.12: Startup Entry into Novel Products, by Soviet S&E Immigration

	DV: First Laser in Class		
	(1) All	(2) Startup	(3) Incumbent
Year >1990 × Soviet Paper Share × Soviet Migrants	1.041	1.565	1.175
	(0.645)	(0.346)	(0.649)
Year >1990 × Soviet Paper Share	0.429	0.621	0.307
	(0.200)	(0.371)	(0.309)
Year >1990 × Soviet Migrants	-0.074	-0.705	0.060
	(0.403)	(0.342)	(0.467)
Soviet Paper Share × Soviet Migrants	-1.051	-1.439	-1.207
	(0.649)	(0.323)	(0.655)
Soviet Paper Share	0.141	-0.049	0.270
	(0.154)	(0.142)	(0.282)
Soviet Migrants	0.131	0.829	-0.012
	(0.432)	(0.357)	(0.509)
China&India Migrants	0.018	0.013	0.020
	(0.005)	(0.013)	(0.008)
ROW Migrants	-0.009	-0.013	-0.009
	(0.004)	(0.011)	(0.006)
Employees	0.000	0.001	0.000
	(0.000)	(0.001)	(0.000)
Average of DV	0.075	0.072	0.076
Year FEs	Yes	Yes	Yes
MSA FEs	Yes	Yes	Yes
R^2	0.574	0.588	0.638
Number of Observations	2,335	746	1,588

Notes: Unit of analysis is the product. *Soviet Migrants* refers to the cumulative number of S&E migrants from the Soviet Union that arrive at the focal firm's MSA-year (in hundreds). Standard errors are clustered at the MSA level.

Bibliography

Acs, Zoltan J and David B Audretsch (1987) "Innovation, market structure, and firm size," *The Review of Economics and Statistics*, 69 (4), 567–574.

Adner, Ron (2002) "When are technologies disruptive? A demand-based view of the emergence of competition," *Strategic Management Journal*, 23 (8), 667–688.

Agrawal, Ajay, Iain Cockburn, and Carlos Rosell (2010) "Not invented here? Innovation in company towns," *Journal of Urban Economics*, 67 (1), 78–89.

Agrawal, Ajay and Rebecca Henderson (2002) "Putting patents in context: Exploring knowledge transfer from MIT," *Management Science*, 48 (1), 44–60.

Ahmadpoor, Mohammad and Benjamin F Jones (2017) "The dual frontier: Patented inventions and prior scientific advance," *Science*, 357 (6351), 583–587.

Ahuja, Gautam and Curba Morris Lampert (2001) "Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions," *Strategic Management Journal*, 22 (6-7), 521–543.

Anderson, Simon P, Andre De Palma, and Jacques-Francois Thisse (1992) *Discrete choice theory of product differentiation*, Cambridge, Mass.: MIT Press.

Anton, James J and Dennis A Yao (1994) "Expropriation and inventions: Appropriable rents in the absence of property rights," *American Economic Review*, 84 (1), 190–209.

Arora, Ashish (1995) "Licensing tacit knowledge: intellectual property rights and the market for know-how," *Economics of innovation and new technology*, 4 (1), 41–60.

Arora, Ashish, Sharon Belenzon, Wesley Cohen, and Honggi Lee (2018a) "Are ideas getting harder to find, or have firms just become too big?" Working Paper.

Arora, Ashish, Sharon Belenzon, Konstantin Kosenko, Jungkyu Suh, and Yishay Yafeh (2021a) "The Rise of Scientific Research in Corporate America," NBER Working Paper No. 29260, National Bureau of Economic Research.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi (2018b) "The decline of science in corporate R&D," *Strategic Management Journal*, 39 (1), 3–32.

——— (2019a) "A theory of the US innovation ecosystem: evolution and the social value of diversity," *Industrial and Corporate Change*, 28 (2), 289–307.

Arora, Ashish, Sharon Belenzon, Andrea Patacconi, and Jungkyu Suh (2019b) "Why the US innovation ecosystem is slowing down," *Harvard Business Review Online*.

——— (2020) "The changing structure of American innovation: Some cautionary remarks for economic growth," *Innovation Policy and the Economy*, 20 (1), 39–93.

Arora, Ashish, Sharon Belenzon, and Lia Sheer (2021b) "Knowledge spillovers and corporate investment in scientific research," *American Economic Review*, 111 (3), 871–98.

Arora, Ashish, Sharon Belenzon, and Jungkyu Suh (2022a) "Science and the Market for Technology," *Management Science*.

Arora, Ashish and Wesley Cohen (2018) "Firm Size, Innovation and the Effectiveness of DoD Support to Corporate R&D," Working Paper.

Arora, Ashish and Andrea Fosfuri (2000) "The market for technology in the chemical industry: causes and consequences," *Revue d'économie industrielle*, 92 (2), 317–334.

Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella (2004) *Markets for technology: The economics of innovation and corporate strategy*, Cambridge, Mass.: MIT Press.

Arora, Ashish, Andrea Fosfuri, and Thomas Roende (2022b) "Caught In The Middle: The Bias Against Startup Innovation With Technical And Commercial Challenges," NBER Working Paper No. 29654, National Bureau of Economic Research.

Arora, Ashish and Alfonso Gambardella (1994) "The changing technology of technological change: general and abstract knowledge and the division of innovative labour," *Research Policy*, 23 (5), 523–532.

Arque-Castells, Pere and Daniel F Spulber (2017) "The Market for Technology: Harnessing Creative Destruction," Working Paper.

Arrow, Kenneth (1962) "Economic welfare and the allocation of resources for invention," in *The rate and direction of inventive activity: Economic and social factors*, 609–626, Princeton: Princeton University Press.

Arts, Sam, Jianan Hou, and Juan Carlos Gomez (2020) "Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures," *Research Policy*, 50 (2), 104–144.

AUTM (2015) "Licensing Activity Survey, FY2015," *Association of University Technology Managers. AUTM.net*.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen (2020) "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 135 (2), 645–709, 10.1093/qje/qjaa004.

Azoulay, Pierre (2002) "Do pharmaceutical sales respond to scientific evidence?" *Journal of Economics & Management Strategy*, 11 (4), 551–594.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin (2015) "Does science advance one funeral at a time?" NBER Working Paper No. 21788, National Bureau of Economic Research.

Azoulay, Pierre, Erica Fuchs, Anna P Goldstein, and Michael Kearney (2019) "Funding breakthrough research: promises and challenges of the "ARPA Model"," *Innovation Policy and the Economy*, 19 (1), 69–96.

Azoulay, Pierre, Joshua S Graff Zivin, Danielle Li, and Bhaven N Sampat (2018) "Public R&D investments and private-sector patenting: evidence from NIH funding rules," *The Review of Economic Studies*, 86 (1), 117–152.

Bagayev, Sergei, Oleg Krokhin, and Alexander Manenkov (2005) "Pages from the history of quantum electronics research in the Soviet Union," *Journal of Modern Optics*, 52 (12), 1657–1669.

Basov, NG and AH Guenther (1985) "History of the development of lasers at the PN Lebedev Physics Institute, Academy of Sciences of the USSR," *Soviet Journal of Quantum Electronics*, 15 (3), 301.

Basting, Dirk, Klaus D Pippert, and Uwe Stamm (2002) "History and future prospects of excimer lasers," in *Second International Symposium on Laser Precision Microfabrication*, 4426, 25–34: International Society for Optics and Photonics.

Bayus, Barry L and Rajshree Agarwal (2007) "The role of pre-entry experience, entry timing, and product technology strategies in explaining firm survival," *Management Science*, 53 (12), 1887–1902.

Belenzon, Sharon and Larisa C Cioaca (2021) "Guaranteed Markets and Corporate Scientific Research," NBER Working Paper No. 28644, National Bureau of Economic Research.

Belenzon, Sharon and Mark Schankerman (2009) "University knowledge transfer: private ownership, incentives, and local development objectives," *The Journal of Law and Economics*, 52 (1), 111–144.

Benner, Mary J and Michael Tushman (2002) "Process management and technological innovation: A longitudinal study of the photography and paint industries," *Administrative Science Quarterly*, 47 (4), 676–707.

Berle, Adolph and Gardiner Means (1932) "Private property and the modern corporation," *New York: Mac-millan*.

Bhardwaj, Gaurab, John C Camillus, and David A Hounshell (2006) "Continual corporate entrepreneurial search for long-term growth," *Management Science*, 52 (2), 248–261.

Bhaskarabhatla, Ajay and Deepak Hegde (2014) "An organizational perspective on patenting and open innovation," *Organization Science*, 25 (6), 1744–1763.

Bhaskarabhatla, Ajay and Steven Klepper (2014) "Latent submarket dynamics and industry evolution: lessons from the US laser industry," *Industrial and Corporate Change*, 23 (6), 1381–1415.

Bikard, M. (2015) "Peer-Based Knowledge Validation: A Hurdle to the Flow of Academic Science to Inventors," Working Paper.

Block, Fred and Matthew R Keller (2009) "Where do innovations come from? Transformations in the US economy, 1970–2006," *Socio-Economic Review*, 7 (3), 459–483.

Bloom, N., M. Schankerman, and J. Van Reenen (2013) "Identifying technology spillovers and product market rivalry," *Econometrica*, 81 (4), 1347–1393.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb (2017) "Are ideas getting harder to find?" NBER Working Paper No. 23782, National Bureau of Economic Research.

Bonbright, James Cummings and Gardiner Coit Means (1932) *The holding company: Its public significance and its regulation*: McGraw-Hill Book Company.

Borjas, George J and Kirk B Doran (2012) "The collapse of the Soviet Union and the productivity of American mathematicians," *The Quarterly Journal of Economics*, 127 (3), 1143–1203.

Boroush, Mark (2017) "National Patterns of R&D Resources: 2014–15 Data Update," NSF 17-311, National Science Foundation, Arlington, VA, https://www.nsf.gov/statistics/2017/nsf17311/.

Bresnahan, T.F. and A. Gambardella (1998) "The division of inventive labor and the extent of the market," in Helpman, Elhanan ed. *General purpose technologies and economic growth*, 253–281, Cambridge, Mass.: MIT Press.

Bromberg, Joan Lisa (1991) *The laser in America, 1950-1970*: Cambridge, Mass. : MIT Press, c1991.

Buchanan, Norman S (1936) "The public utility holding company problem," *Calif. L. Rev.*, 25, 517.

Bush, Vannevar (1945) "Science: The endless frontier," *Transactions of the Kansas Academy of Science (1903-)*, 48 (3), 231–264.

Carnegie, Andrew (2006) *The autobiography of Andrew Carnegie and the gospel of wealth*, New York: Signet.

Castelvecchi, Davide (2016) "Majority of mathematicians hail from just 24 scientific 'families'," *Nature News*, 537 (7618), 20–21.

Cattani, Gino (2005) "Preadaptation, firm heterogeneity, and technological performance: A study on the evolution of fiber optics, 1970–1995," *Organization Science*, 16 (6), 563–580.

Chandler Jr, Alfred D (1977) *The visible hand*, Cambridge, Mass.: Belknap Press.

Chang, Sea-Jin and Brian Wu (2014) "Institutional barriers and industry dynamics," *Strategic Management Journal*, 35 (8), 1103–1123.

Chatterjee, Sayan and Birger Wernerfelt (1991) "The link between resources and type of diversification: Theory and evidence," *Strategic Management Journal*, 12 (1), 33–48.

Chatterji, Aaron K (2009) "Spawned with a silver spoon? Entrepreneurial performance and innovation in the medical device industry," *Strategic Management Journal*, 30 (2), 185–206.

Christensen, Clayton M (1993) "The rigid disk drive industry: A history of commercial and technological turbulence," *Business History Review*, 67 (4), 531–588.

Cohen, Wesley M. (2010) "Fifty Years of Empirical Studies of Innovative Activity and Performance," in Hawley, Teresa S. and Robert G. Hawley eds. *Handbook of the Economics of Innovation, Volume 1*, 129–213, Amsterdam; Boston: Elsevier.

Cohen, Wesley M. and Steven Klepper (1992) "The tradeoff between firm size and diversity in the pursuit of technological progress," *Small Business Economics*, 4 (1), 1–14.

Cohen, Wesley M and Steven Klepper (1996) "Firm size and the nature of innovation within industries: the case of process and product R&D," *The Review of Economics and Statistics*, 78 (2), 232–243.

Cohen, Wesley M and Daniel A Levinthal (1989) "Innovation and learning: the two faces of R & D," *The economic journal*, 99 (397), 569–596.

Cotropia, Lemley Mark A., Christopher Anthony and Bhaven N. Sampat (2013) "Do Applicant Patent Citations Matter?" *Research Policy*, 42 (4), 844–854.

Council, National Research (1938) *Research: A National Resource : II. Industrial Research*, Washington, DC: The National Academies Press, https://www.nap.edu/catalog/20233/research-a-national-resource-ii-industrial-research.

Cunningham, Colleen, Florian Ederer, and Song Ma (2021) "Killer acquisitions," *Journal of Political Economy*, 129 (3), 649–702.

De Marco, Antonio, Giuseppe Scellato, Elisa Ughetto, and Federico Caviggioli (2017) "Global markets for technology: Evidence from patent transactions," *Research Policy*, 46 (9), 1644–1654.

Dernis, Hélène and Mosahid Khan (2004) "Triadic patent families methodology," OECD Science, Technology and Industry Working Papers , OECD, https://www.oecd-ilibrary.org/science-and-technology/triadic-patent-families-methodology_443844125004.

Eggers, JP (2014) "Competing technologies and industry evolution: The benefits of making mistakes in the flat panel display industry," *Strategic Management Journal*, 35 (2), 159–178.

Eggers, JP and K Francis Park (2018) "Incumbent adaptation to technological change: The past, present, and future of research on heterogeneous incumbent response," *Academy of Management Annals*, 12 (1), 357–389.

Evenson, Robert E and Yoav Kislev (1976) "A stochastic model of applied research," *Journal of Political Economy*, 84 (2), 265–281.

Ewens, Michael and Christian Fons-Rosen (2013) "The consequences of entrepreneurial firm founding on innovation," Working Paper, https://ssrn.com/abstract=2291811.

Ferrucci, Edoardo (2020) "Migration, innovation and technological diversion: German patenting after the collapse of the Soviet Union," *Research Policy*, 49 (9), 104057.

Field, Alexander J (2003) "The most technologically progressive decade of the century," *American Economic Review*, 93 (4), 1399–1413.

Figueroa, Nicolás and Carlos J Serrano (2019) "Patent trading flows of small and large firms," *Research Policy*, 48 (7), 1601–1616.

Fleming, Donald and Bernard Bailyn (1969) *The intellectual migration: Europe and America, 1930-1960*, 2, Cambridge, Mass.: Belknap Press of Harvard University Press.

Fleming, L., H. Greene, G. Li, M. Marx, and D. Yao (2019) "Government-funded research increasingly fuels innovation," *Science*, 364 (6446), 1139.

Fleming, Lee (2001) "Recombinant uncertainty in technological search," *Management Science*, 47 (1), 117–132.

Fleming, Lee and Olav Sorenson (2004) "Science as a map in technological search," *Strategic Management Journal*, 25 (8-9), 909–928.

Franco, April M, MB Sarkar, Rajshree Agarwal, and Raj Echambadi (2009) "Swift and smart: The moderating effects of technological capabilities on the market pioneering–firm survival relationship," *Management Science*, 55 (11), 1842–1860.

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

Galasso, Alberto and Mark Schankerman (2014) "Patents and cumulative innovation: Causal evidence from the courts," *The Quarterly Journal of Economics*, 130 (1), 317–369.

——— (2018) "Patent rights, innovation, and firm exit," *The RAND Journal of Economics*, 49 (1), 64–86.

Galasso, Alberto, Mark Schankerman, and Carlos J Serrano (2013) "Trading and enforcing patent rights," *The RAND Journal of Economics*, 44 (2), 275–312.

Ganco, Martin and Rajshree Agarwal (2009) "Performance differentials between diversifying entrants and entrepreneurial start-ups: A complexity approach," *Academy of Management Review*, 34 (2), 228–252.

Ganguli, Ina (2015) "Immigration and Ideas: What Did Russian Scientists "Bring" to the United States?" *Journal of Labor Economics*, 33 (S1), S257–S288.

Gans, Joshua S., David H. Hsu, and Scott Stern (2002) "When Does Start-Up Innovation Spur the Gale of Creative Destruction?" *The RAND Journal of Economics*, 33 (4), 571–586.

Gans, Joshua S, David H Hsu, and Scott Stern (2008) "The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays," *Management Science*, 54 (5), 982–997.

Gans, Joshua S, Michael Kearney, Erin L Scott, and Scott Stern (2020) "Choosing Technology: An Entrepreneurial Strategy Approach," NBER Working Paper No. 27489, National Bureau of Economic Research.

Gans, Joshua S. and Scott Stern (2010) "Is there a market for ideas?" *Industrial and Corporate Change*, 19 (3), 805–837.

Gatignon, Hubert, Michael L Tushman, Wendy Smith, and Philip Anderson (2002) "A structural approach to assessing innovation: Construct development of innovation locus, type, and characteristics," *Management Science*, 48 (9), 1103–1122.

Geiger, Roger L. (1986) *To advance knowledge : the growth of American research universities, 1900-1940*: New York : Oxford University Press, 1986.

Gilbert, Clark G (2005) "Unbundling the structure of inertia: Resource versus routine rigidity," *Academy of Management Journal*, 48 (5), 741–763.

Gilbert, Richard (2006) "Looking for Mr. Schumpeter: Where Are We in the Competition–Innovation Debate?" *Innovation Policy and the Economy*, 6, 159–215.

Gilbert, Richard J and David MG Newbery (1982) "Preemptive patenting and the persistence of monopoly," *American Economic Review*, 514–526.

Gordon, Robert J (2012) "Is US economic growth over? Faltering innovation confronts the six headwinds," NBER Working Paper No. 18315, National Bureau of Economic Research.

Gordon, Robert J. (2016) *The rise and fall of American growth : the U.S. standard of living since the Civil War*, Princeton, New Jersey ; Oxford, England: Princeton University Press.

Graham, John R, Mark T Leary, and Michael R Roberts (2015) "A century of capital structure: The leveraging of corporate America," *Journal of Financial Economics*, 118 (3), 658–683.

Graham, Loren R (1993) *Science in Russia and the Soviet Union: A short history*, Cambridge ; New York: Cambridge University Press.

Graham, Stuart JH, Alan C Marco, and Amanda F Myers (2018) "Patent transactions in the marketplace: Lessons from the uspto patent assignment dataset," *Journal of Economics & Management Strategy*, 27 (3), 343–371.

Griliches, Zvi (1957) "Hybrid corn: An exploration in the economics of technological change," *Econometrica, Journal of the Econometric Society*, 501–522.

——— (1990) "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28 (4), 1661–1707.

Gross, Daniel P and Bhaven N Sampat (2020a) "Inventing the Endless Frontier: The Effects of the World War II Research Effort on Post-War Innovation," harvard business school strategy unit working paper (20-126).

——— (2020b) "Organizing Crisis Innovation: Lessons from World War II," NBER Working Paper No. 27909, National Bureau of Economic Research.

Gutiérrez, Germán and Thomas Philippon (2020) "Some facts about dominant firms," NBER Working Paper No. 27985, National Bureau of Economic Research.

Hall, Bronwyn H and Josh Lerner (2010) "The financing of R&D and innovation," in Hawley, Teresa S. and Robert G. Hawley eds. *Handbook of the Economics of Innovation, Volume 1*, 1, 609–639, Amsterdam; Boston: Elsevier.

Haynes, Williams (1945) *American Chemical Industry: The World War I period: 1912-1922*, 3, New York: Van Nostrand.

Hecht, Jeff (1999) *Laser Guidebook*, New York: McGraw-Hill.

Helfat, Constance E (1997) "Know-how and asset complementarity and dynamic capability accumulation: the case of R&D," *Strategic Management Journal*, 18 (5), 339–360.

Helfat, Constance E and Marvin B Lieberman (2002) "The birth of capabilities: market entry and the importance of pre-history," *Industrial and Corporate Change*, 11 (4), 725–760.

Henderson, Rebecca (1993) "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry," *The RAND Journal of Economics*, 24 (2), 248–270.

Henderson, Rebecca M (1992) "Technological change and the management of architectural knowledge," in Kochan, Thomas and Michael Useem eds. *Transforming organizations*, 118–131, New York: Oxford University Press.

Henderson, Rebecca M and Kim B Clark (1990) "Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms," *Administrative Science Quarterly*, 35 (1), 9–30.

Hicks, Diana (1995) "Published papers, tacit competencies and corporate management of the public/private character of knowledge," *Industrial and corporate change*, 4 (2), 401–424.

Hollings, Christopher D (2016) *Scientific communication across the Iron Curtain*: Springer.

Holton, Gerald (1981) "The formation of the American physics Community in the 1920s and the coming of Albert Einstein," *Minerva*, 19 (4), 569–581.

Hounshell, David A (1990) "Interpreting the History of Industrial Research and Development: The Case of EI du Pont de Nemours & Co.," *Proceedings of the American Philosophical Society*, 134 (4), 387–407.

——— (1998) "Measuring the Return on Investment in R&D: Voices from the Past, Visions of the Future," in *Assessing the Value of Research in the Chemical Sciences: Report of a Workshop*, 6–14, Washington DC: National Academy Press, http://www.nap.edu/catalog/6200.html.

Hounshell, David and John Kenly Jr. Smith (1988) *Science and corporate strategy : Du Pont R&D, 1902-1980*: Cambridge Cambridgeshire ; New York : Cambridge University Press, 1988.

Howell, Sabrina T (2017) "Financing innovation: Evidence from R&D grants," *American Economic Review*, 107 (4), 1136–64.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2018) "Frontier knowledge and scientific production: Evidence from the collapse of international science," *The Quarterly Journal of Economics*, 133 (2), 927–991.

Jayaraj, Sebastian and Michelle Gittelman (2018) "Scientific maps and innovation: Impact of the human genome on drug discovery," in *DRUID Society Conference*, 1–56.

Jenkins, Reese and Alfred Dupont Chandler (1975) *Images and enterprise: Technology and the American photographic industry, 1839 to 1925*, Baltimore: Johns Hopkins University Press.

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

Kamien, Morton I and Nancy L Schwartz (1982) *Market structure and innovation*: Cambridge University Press.

Kandel, Eugene, Konstantin Kosenko, Randall Morck, and Yishay Yafeh (2019) "The Great Pyramids of America: A Revised History of US Business Groups, Corporate Ownership and Regulation, 1926-1950," *Strategic Management Journal*, 40 (5), 781–808.

Kao, Jennifer (2019) *Essays in the Economics of Health and Innovation* PhD Thesis.

Kapoor, Rahul and Nathan R Furr (2015) "Complementarities and competition: Unpacking the drivers of entrants' technology choices in the solar photovoltaic industry," *Strategic Management Journal*, 36 (3), 416–436.

Karlov, NV, ON Krokhin, and SG Lukishova (2010) "History of quantum electronics at the Moscow Lebedev and General Physics Institutes: Nikolaj Basov and Alexander Prokhorov," *Applied Optics*, 49 (25), F32–F46.

Kerr, William R and Ramana Nanda (2015) "Financing innovation," *Annual Review of Financial Economics*, 7, 445–462.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf (2014) "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, 28 (3), 25–48.

Kesavan, Subramanian (1994) *Rheology of aqueous gels* PhD Thesis, Princeton University.

Kevles, Daniel (1979) "The physics, mathematics, and chemistry communities: A comparative analysis," in Oleson, Alexandra and John Voss eds. *The Organization of Knowledge in Modern America, 1860-1920*, 139–172, Baltimore: Johns Hopkins University Press.

Khanna, Tarun (2000) "Business groups and social welfare in emerging markets: Existing evidence and unanswered questions," *European Economic Review*, 44 (4-6), 748–761.

——— (2018) "When technology gets ahead of society," *Harvard Business Review*, 96 (4), 86–95.

Khanna, Tarun and Krishna Palepu (2000) "Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups," *The Journal of Finance*, 55 (2), 867–891.

King, Andrew A and Christopher L Tucci (2002) "Incumbent entry into new market niches: The role of experience and managerial choice in the creation of dynamic capabilities," *Management Science*, 48 (2), 171–186.

Klapper, Leora, Luc Laeven, and Raghuram Rajan (2006) "Entry regulation as a barrier to entrepreneurship," *Journal of financial economics*, 82 (3), 591–629.

Klepper, Steven (1996) "Entry, exit, growth, and innovation over the product life cycle," *American Economic Review*, 86 (3), 562–583.

——— (2007) "Disagreements, spinoffs, and the evolution of Detroit as the capital of the US automobile industry," *Management Science*, 53 (4), 616–631.

——— (2015) *Experimental capitalism: the nanoeconomics of American high-tech industries*, Princeton, New Jersey: Princeton University Press.

Klepper, Steven and Kenneth L Simons (2000) "Dominance by birthright: entry of prior radio producers and competitive ramifications in the US television receiver industry," *Strategic Management Journal*, 21 (10-11), 997–1016.

Klepper, Steven and Sally Sleeper (2005) "Entry by spinoffs," *Management Science*, 51 (8), 1291–1306.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman (2017) "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 132 (2), 665–712.

Kogut, Bruce and Udo Zander (1992) "Knowledge of the firm, combinative capabilities, and the replication of technology," *Organization science*, 3 (3), 383–397.

Krieger, Joshua L., Danielle Li, and Dimitris Papanikolaou (2018) "Missing Novelty in Drug Development," NBER Working Paper No. 24595, National Bureau of Economic Research.

Kuhn, Jeffrey M (2016) "Property rights and frictions in the sale of patents."

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

Kuznets, S. (1971) *Nobel Prize Lecture: Modern Economic Growth: Findings and Reflections*: Nobelprize.org.

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming (2011) "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)."

Lamoreaux, Naomi R. and Kenneth L. Sokoloff (1997) *Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries*: National Bureau of Economic Research Cambridge, Mass., USA.

Lee, David L, Justin McCrary, Marcelo J Moreira, and Jack Porter (2020) "Valid t-ratio Inference for IV," *arXiv preprint arXiv:2010.05058*.

Lerner, Josh (2000) "Assessing the contribution of venture capital," *the RAND Journal of Economics*, 31 (4), 674–692.

Lerner, Josh and Ramana Nanda (2020) "Venture capital's role in financing innovation: What we know and how much we still need to learn," *Journal of Economic Perspectives*, 34 (3), 237–61.

Levin, Richard C, Alvin K Klevorick, Richard R Nelson, Sidney G Winter, Richard Gilbert, and Zvi Griliches (1987) "Appropriating the returns from industrial research and development," *Brookings papers on economic activity*, 1987 (3), 783–831.

Lichtenberg, Frank R (1987) "The effect of government funding on private industrial research and development: a re-assessment," *The Journal of Industrial Economics*, 36 (1), 97–104.

Link, Albert N (1982) "A disaggregated analysis of industrial R&D: Product versus process innovation," in Devendra, Sahal ed. *The transfer and utilization of technical knowledge*, 45–62, Lexington, Mass.: Lexington Books.

Lourdes Sosa, M (2013) "Decoupling market incumbency from organizational prehistory: Locating the real sources of competitive advantage in R&D for radical innovation," *Strategic Management Journal*, 34 (2), 245–255.

Lundberg, Ferdinand (1937) *America's 60 families*, New York: Vanguard Press.

Ma, Song, Wei Wang, and others (2017) "Selling innovation in bankruptcy," Working Paper.

Maclaurin, William Rupert and R Joyce Harman (1949) *Invention & innovation in the radio industry*, New York: Macmillian.

MacLeod, W Bentley and Miguel Urquiola (2021) "Why Does the United States Have the Best Research Universities? Incentives, Resources, and Virtuous Circles," *Journal of Economic Perspectives*, 35 (1), 185–206.

Manadhata, Pratyusa K, Sandeep Yadav, Prasad Rao, and William Horne (2014) "Detecting malicious domains via graph inference," in *European Symposium on Research in Computer Security*, 1–18: Springer.

Mansfield, E. (1980) "Basic research and productivity increase in manufacturing," *American Economic Review*, 70 (5), 863–873.

Mansfield, Edwin (1981) "Composition of R and D expenditures: Relationship to size of firm, concentration, and innovative output," *The Review of Economics and Statistics*, 63 (4), 610–615.

Marco, Alan C, Amanda F Myers, Stuart JH Graham, Paul A D'Agostino, and Kirsten Apple (2015) "The USPTO patent assignment dataset: Descriptions and analysis," USPTO Economic working paper no. 2015-2, Office of Chief Economist, US Patent and Trademark Office, Alexandria, VA.

Martens, John A. (2020) "Secrecy in teh USSR and German Democratic Republic Patent Systems."

Marx, Matt and Aaron Fuegi (2020) "Reliance on science: Worldwide front-page patent citations to scientific articles," *Strategic Management Journal*, 41 (9), 1572–1594.

Mazzucato, Mariana (2015) *The entrepreneurial state : debunking public vs. private sector myths*, New York, NY: PublicAffairs.

McMillan, Frank M (1979) *The chain straighteners: Fruitful innovation: the discovery of linear and stereoregular synthetic polymers*, London: Springer.

Merges, Robert P and Richard R Nelson (1990) "On the complex economics of patent scope," *Columbia Law Review*, 90 (4), 839–916.

Merrill, Stephen A. (2018) "Righting the Research Imbalance,", The Center for Innovation Policy at Duke Law.

Merton, Robert K (1957) "Priorities in scientific discovery: a chapter in the sociology of science," *American Sociological Review*, 22 (6), 635–659.

Mirzabekov, Andrei D (1993) "Visit Russia, help science," *Nature*, 363 (6430), 576–576.

Mitchell, Will (1989) "Whether and When? Probability and Timing of Incumbents' Entry into Emerging Industrial Subfields," *Administrative Science Quarterly*, 34 (2), 208–230.

Mody, Cyrus CM (2011) *Instrumental community: Probe microscopy and the path to nanotechnology*, Cambridge, Mass.: MIT Press.

Moeen, Mahka and Rajshree Agarwal (2017) "Incubation of an industry: Heterogeneous knowledge bases and modes of value capture," *Strategic Management Journal*, 38 (3), 566–587.

Moeen, Mahka and Will Mitchell (2020) "How do pre-entrants to the industry incubation stage choose between alliances and acquisitions for technical capabilities and specialized complementary assets?" *Strategic Management Journal*, 41 (8), 1450–1489.

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

Montgomery, Cynthia A and Sam Hariharan (1991) "Diversified expansion by large established firms," *Journal of Economic Behavior & Organization*, 15 (1), 71–89.

Moretti, Enrico, Claudia Steinwender, and John Van Reenen (2019) "The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers," NBER Working Paper No. 26483, National Bureau of Economic Research.

Moser, Petra and Shmuel San (2020) "Immigration, Science, and Invention. Lessons from the Quota Acts, Lessons from the Quota Acts," Working Paper.

Moser, Petra, Alessandra Voena, and Fabian Waldinger (2014) "German Jewish émigrés and US invention," *American Economic Review*, 104 (10), 3222–55.

Mowery, David C (1983) "The relationship between intrafirm and contractual forms of industrial research in American manufacturing, 1900–1940," *Explorations in economic history*, 20 (4), 351–374.

——— (1998) "The changing structure of the US national innovation system: implications for international conflict and cooperation in R&D policy," *Research Policy*, 27 (6), 639–654.

Mowery, David C. (2009) "Plus ca change: Industrial R&D in the "third industrial revolution"," *Industrial and corporate change*, 18 (1), 1–50.

Mowery, David and Nathan Rosenberg (1998) *Paths of innovation : technological change in 20th-century America*: Cambridge ; New York : Cambridge University Press, 1998.

Narin, Francis, Kimberly S Hamilton, and Dominic Olivastro (1997) "The increasing linkage between US technology and public science," *Research Policy*, 26 (3), 317–330.

Natta, Giulio, Piero Pino, Paolo Corradini, Ferdinando Danusso, Enrico Mantica, Giorgio Mazzanti, and Giovanni Moraglio (1955) "Crystalline High Polymers of -Olefins," *Journal of the American Chemical Society*, 77 (6), 1708–1710.

Nelson, Richard R (1959a) "The economics of invention: A survey of the literature," *The Journal of Business*, 32 (2), 101–127.

——— (1959b) "The simple economics of basic scientific research," *Journal of Political Economy*, 67 (3), 297–306.

——— (1981) "Assessing private enterprise: an exegesis of tangled doctrine," *The Bell Journal of Economics*, 12 (1), 93–111.

——— (1982) "The role of knowledge in R&D efficiency," *Quarterly Journal of Economics*, 97 (3), 453–470.

Nicholas, Tom (2010) "The role of independent invention in US technological development, 1880–1930," *The Journal of Economic History*, 70 (1), 57–82.

——— (2011) "Did R&D firms used to patent? Evidence from the first innovation surveys," *The Journal of Economic History*, 71 (4), 1032–1059.

——— (2014) "Scale and innovation during two US breakthrough eras," Harvard Business School Working Paper, No. 15-038, Harvard Business School.

Nutter, G Warren and Henry Adler Einhorn (1969) *Enterprise monopoly in the United States: 1899-1958*, New York: Columbia University Press.

Pavitt, Keith (1987) "The objectives of technology policy," *Science and public policy*, 14 (4), 182–188.

Pece, Christopher (2016) "Putting the Cart Before a Lame Horse: A case study for future initiatives to automate the use of administrative records for reporting government R&D,", National Center for Science and Engineering Statistics, National Science Foundation.

Perazich, George and Philip M Field (1940) *Industrial research and changing technology* (4): WPA, Nat. Research Project.

Perkmann, Markus, Rossella Salandra, Valentina Tartari, Maureen McKelvey, and Alan Hughes (2021) "Academic engagement: A review of the literature 2011-2019," *Research Policy*, 50 (1), 104114.

Pinchot, Gifford (1928) *The power monopoly; its make-up and its menace*, Milford, Pa: McFetridge.

Png, Ivan PL (2019) "US R&D, 1975–1998: A new dataset," *Strategic Management Journal*, 40 (5), 715–735.

Polanyi, Michael (1958) *Personal knowledge: Towards a post-critical philosophy*, Chicago: University of Chicago Press.

Prokhorov, Alexander M and Ivan Alexandrov Shcherbakov (1991) "Soviet developments in solid state lasers," in *Solid State Lasers II*, 1410, 70–88: International Society for Optics and Photonics.

Prusa, Thomas J and James A Schmitz Jr (1991) "Are new firms an important source of innovation?: Evidence from the PC software industry," *Economics Letters*, 35 (3), 339–342.

Reich, Leonard S. (1983) "Irving Langmuir and the Pursuit of Science and Technology in the Corporate Environment," *Technology and Culture*, 24 (2), 199–221.

Reich, Leonard S (1985) *The making of American industrial research : science and business at GE and Bell, 1876-1926*, Cambridge Cambridgeshire ; New York: Cambridge University Press.

Reinganum, Jennifer F (1983) "Uncertain innovation and the persistence of monopoly," *American Economic Review*, 73 (4), 741–748.

Richter, Ralf and Jochen Streb (2011) "Catching-up and falling behind: knowledge spillover from American to German machine toolmakers," *The Journal of Economic History*, 71 (4), 1006–1031.

Roach, Michael and Wesley M Cohen (2013) "Lens or prism? Patent citations as a measure of knowledge flows from public research," *Management Science*, 59 (2), 504–525.

Robbins, Carol A (2009) "Measuring payments for the supply and use of intellectual property," in Reinsdorf, Marshall and Matthew Slaughter eds. *International trade in services and intangibles in the era of globalization*, 139–171, Chicago: University of Chicago Press.

Rosenberg, Nathan (1974) "Science, Invention and Economic Growth," *The Economic Journal*, 84 (333), 90–108.

——— (1994) *Exploring the black box: Technology, economics, and history*, Cambridge, England; New York: Cambridge University Press.

Rotemberg, Julio J and Garth Saloner (1994) "Benefits of narrow business strategies," *American Economic Review*, 84 (5), 1330–1349.

Rothaermel, Frank T, Shanti D Agung, and Lin Jiang (2007) "University entrepreneurship: a taxonomy of the literature," *Industrial and Corporate Change*, 16 (4), 691–791.

Rothaermel, Frank T and Charles WL Hill (2005) "Technological discontinuities and complementary assets: A longitudinal study of industry and firm performance," *Organization Science*, 16 (1), 52–70.

Rüegg, Walter (2004) *A history of the university in Europe: Volume 3, universities in the nineteenth and early twentieth centuries (1800–1945)*, 3, Cambridge, UK; New York: Cambridge University Press.

Sah, R.J. and J.E. Stiglitz (1986) "The architecture of economic systems: Hierarchies and Polyarchies," *American Economic Review*, 76 (4), 716–727.

Sauermann, H. and W. M. Cohen (2010) "What makes them tick? Employee motives and firm innovation," *Management Science*, 56 (12), 2134–2153.

Schawlow, Arthur L and Charles H Townes (1958) "Infrared and optical masers," *Physical Review*, 112 (6), 1940–1949.

Schmookler, Jacob (1962) "Economic sources of inventive activity," *The Journal of Economic History*, 22 (1), 1–20.

Schumpeter, Joseph A., 1883-1950 (1950) *Capitalism, Socialism, and Democracy.*, New York: New York, Harper.

Schumpeter, Joseph Alois (1939) *Business cycles*, 1: McGraw-Hill New York.

Serrano, Carlos J (2010) "The dynamics of the transfer and renewal of patents," *The RAND Journal of Economics*, 41 (4), 686–708.

Serrano, Carlos J. (2018) "Estimating the Gains from Trade in the Market for Patent Rights," *International Economic Review*, 59 (4), 1877–1904.

Service, National Research Council (US) Research Information (1931) *Industrial Research Laboratories of the United States, Including Consulting Research Laboratories*: RR Bowker Company.

Servos, John W (1994) "Changing partners: The Mellon Institute, private industry, and the federal patron," *Technology and Culture*, 35 (2), 221–257.

Sethi, A.K. (2013) *The Business of Electronics: A Concise History*, New York, NY: Palgrave Macmillan US.

Shimizu, Hiroshi (2019) *General Purpose Technology, Spin-Out, and Innovation: Technological Development of Laser Diodes in the United States and Japan*, Advances in Japanese Business and Economics, Singapore: Springer.

Skiti, Tedi (2020) "Institutional entry barriers and spatial technology diffusion: Evidence from the broadband industry," *Strategic Management Journal*, 41 (7), 1336–1361.

Sleeper, Sally D (1999) *The role of firm capabilities in the evolution of the laser industry: The making of a high-tech market.* Ph.D. dissertation, Carnegie Mellon University.

Smith, John K and David A Hounshell (1985) "Wallace H. Carothers and fundamental research at Du Pont," *Science*, 229 (4712), 436–442.

Stern, S. (2004) "Do scientists pay to be scientists?" *Management Science*, 50 (6), 835–853.

Stevenson, Richard (2009) "The LED's dark secret," *IEEE Spectrum*, 46 (8), 26–31.

Stigler, George J. (1949) "Competition in the United States," in *Five Lectures on Economic Problems*, pp.46–62, London: Longmans, Green and Co.

Sørensen, Jesper B and Toby E Stuart (2000) "Aging, obsolescence, and organizational innovation," *Administrative Science Quarterly*, 45 (1), 81–112.

Teece, David J (1986) "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy," *Research Policy*, 15 (6), 285–305.

Teodoridis, Florenta, Michael Bikard, and Keyvan Vakili (2019) "Creativity at the Knowledge Frontier: The Impact of Specialization in Fast- and Slow-paced Domains," *Administrative Science Quarterly*, 64 (4), 894–927.

Thompson, Peter (2005) "Selection and firm survival: evidence from the shipbuilding industry, 1825–1914," *Review of Economics and Statistics*, 87 (1), 26–36.

Thurgood, Lori, Mary J Golladay, and Susan T Hill (2006) "US Doctorates in the 20th Century," *Arlington, VA: National Science Foundation*.

Tripsas, Mary (1997) "Unraveling the process of creative destruction: Complementary assets and incumbent survival in the typesetter industry," *Strategic Management Journal*, 18 (S1), 119–142.

Tushman, Michael L. and Philip Anderson (1986) "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly*, 31 (3), 439–465.

Urquiola, Miguel (2020) *Markets, Minds, and Money: Why America Leads the World in University Research*, Cambridge, MA: Harvard University Press.

Utterback, James M and William J Abernathy (1975) "A dynamic model of process and product innovation," *Omega*, 3 (6), 639–656.

Veugelers, Reinhilde and Jian Wang (2019) "Scientific novelty and technological impact," *Research Policy*, 48 (6), 1362–1372.

Von Hippel, Eric (1994) ""Sticky information" and the locus of problem solving: implications for innovation," *Management Science*, 40 (4), 429–439.

Weber, Marvin J (2000) *Handbook of lasers*, 18, Boca Raton: CRC Press.

Wilcox, Clair (1940) "Competition and Monopoly in American Industry," *Investigation of Concentration of Economic Power, T.N.E.C, Monograph No. 21*.

Wise, George (1985) *Willis R. Whitney, General Electric, and the origins of US industrial research*, New York: Columbia University Press.

Wu, Brian, Zhixi Wan, and Daniel A Levinthal (2014) "Complementary assets as pipes and prisms: Innovation incentives and trajectory choices," *Strategic Management Journal*, 35 (9), 1257–1278.

Ziedonis, RH (2004) "Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms," *Management Science*, 50 (6), 804–820.

Zucker, Lynne G, Michael R Darby, and Jeff S Armstrong (2002) "Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology," *Management Science*, 48 (1), 138–153.

Biography

Jungkyu Suh was born in Seoul, South Korea. He holds a B.Sc. in International Relations (2010) and an M.Sc. in International Political Economy (2011) from the London School of Economics and Political Science (LSE). He has pursued his Ph.D. in Business Administration at the Fuqua School of Business, Duke University.

Jungkyu was awarded the Aleane Webb Dissertation Research Fellowship and a research grant from the Israeli Science Foundation as a project collaborator. Jungkyu has received the Duke Graduate Student Fellowship throughout his Ph.D. studies. He was also the recipient of the Kwanjeong International Scholarship during his undergraduate studies at the LSE.

Jungkyu will start as an Assistant Professor in the Management and Organizations Department at the Stern School of Business, New York University, in September 2022.