Essays in the Economics of Innovation

Essays in the Economics of Innovation

A dissertation presented

by

Xavier Jaravel

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor:

Professor Raj Chetty

Author:

Xavier Jaravel

Essays in the Economics of Innovation

Abstract

This dissertation examines the social and economic processes that generate innovation and distribute its rewards in society, in the context of the United States over the past twenty years. Chapters 1 and 4 investigate the two-way relationship between innovation and inequality. Chapters 2 and 3 study two recent and important trends in the US innovation system: the rise of teamwork and the activities of patent trolls.

Using detailed product-level data in the retail sector in the United States, Chapter 1 shows that product innovations disproportionately benefited high-income households due to increasing inequality and the endogenous response of supply to market size. From 2004 to 2013, annualized quality-adjusted inflation was 0.65 percentage points lower for high-income households, relative to low-income households. Using national and local changes in market size driven by demographic trends plausibly exogenous to supply factors, this chapter provides causal evidence that a shock to the relative demand for goods (1) affects the direction of product innovations, and (2) leads to a decrease in the relative price of the good for which demand became relatively larger (i.e. the long-term supply curve is downward sloping). A calibration shows that this effect is sufficiently strong to explain most of the observed difference in quality-adjusted inflation rates across the income distribution.

Chapter 2 demonstrates the importance of team-specific capital in the typical inventor's career. Using administrative tax and patent data for the population of US patent inventors

from 1996 to 2012 and the premature deaths of 4,714 inventors, an inventor's premature death is found to cause a large and long-lasting decline in their co-inventor's earnings and citation-weighted patents (-4% and - 15% after 8 years, respectively). Firm disruption, network effects and top-down spillovers are ruled out as primary drivers of this result. Consistent with the team-specific capital interpretation, the effect is larger for more closely-knit teams and primarily applies to co-invention activities.

Chapter 3 investigates the patent acquisition behavior of non-practicing entities (NPEs), also-known as patent trolls. Unlike regular firms, NPEs purchase and assert patents that were granted by a specific set of examiners at the United States Patent Office (USPTO), who tend to allow incremental patents with vaguely-worded claims. The methodology introduced in this chapter leverages the random assignment of patent applications to examiners and provides a novel way of inferring the nature of a patent from prosecution data. A cost-benefit calibration suggests that investments in improving the quality of the examination process at the USPTO would have large social returns.

Using administrative records on the population of individuals who applied for or were granted a patent between 1996 and 2014, Chapter 4 characterizes the lives of more than 1.2 million inventors in the United States. Children of low-income parents are much less likely to become inventors than their higher-income counterparts and decompositions indicate that this income-innovation gap can largely be accounted for by differences in human capital acquisition while children are growing up. The importance of exposure effects during childhood is established by showing that growing up in an area with a high innovation rate in a particular technology class is associated with a much higher probability of becoming an inventor specifically in that technology class. Taken together, these descriptive findings shed light on which types of policy tools are likely to be most effective in sparking innovation. In particular, they suggest that "extensive margin" policies drawing more talented individuals from low-income families into innovation have great potential.

Contents

Abstract **iii**

Acknowledgments **xxii**

1. The Unequal Gains from Product Innovations: Evidence from the Retail Sector **1**

1.1. Introduction 1

1.2. Data Sources 6

1.2.1. Scanner Data 6

1.2.2. Markup Data 8

1.2.3. Manufacturer Data 9

1.3. Measuring Quality-Adjusted Inflation Across the Income Distribution 10

1.3.1. Nonhomothetic Preferences, Product Variety and Real Inequality . . 10

1.3.2. Overview of Methodology and Review of Basic Price Indices 12

1.3.3. Estimation Framework for Nested CES Exact Price Index 15

1.3.4. Inflation Across the Income Distribution For Goods Available in Consecutive Years 19

1.3.5. Changes in Product Variety Across the Income Distribution 28

1.3.6. Quality-Adjusted Inflation Across the Income Distribution 32

1.3.7. Further Robustness Checks . . . 33

1.4. The Equilibrium Response of Supply to Changes in Demand . . . 37

1.4.1. Theory . . . 37

1.4.2. The Relationship Between Inflation and Share of Spending on New Products . . . 43

1.4.3. The Role of Markups . . . 46

1.4.4. The Causal Effect of Growing Demand on Product Innovations and Inflation . . . 48

1.4.5. Alternative Mechanisms . . . 57

1.5. Conclusion . . . 59

2. Team-Specific Capital and Innovation[1] **61**

2.1. Introduction . . . 61

2.2. Data and Descriptive Statistics . . . 66

2.2.1. Data Construction . . . 66

2.2.2. Identifying Deceased Inventors, Survivor Co-inventors, Second-Degree Connections and Coworkers . . . 67

2.2.3. Variable Definitions and Summary Statistics . . . 69

2.2.4. Descriptive Statistics on Patent Inventor Teams . . . 72

2.3. Estimating the Causal Effect of the Premature Death of a Co-Inventor on an Inventor's Compensation and Patents . . . 75

2.3.1. Research Design . . . 75

2.3.2. Regression Framework . . . 78

2.3.3. Results . . . 80

2.3.4. Robustness Checks . . . 84

[1]Co-authored with Neviana Petkova and Alex Bell.

2.4. Distinguishing Between Mechanisms 87
2.4.1. Firm Disruption and Network Effects Are Not the Primary Mechanism 88
2.4.2. Top-Down Spillovers Are Not the Driving Force 92
2.4.3. The Effect Is Driven by Close-Knit Teams 95
2.4.4. Team-Specific Capital as a Likely Mechanism 98
2.5. Conclusion 103

3. Patent Trolls and the Patent Examination Process[2] 105

3.1. Introduction 105
3.2. Background 110
3.2.1. USPTO Examination Process 110
3.2.2. Relationship to Existing Literature 113
3.3. Data and Descriptive Statistics 116
3.3.1. Data Construction 116
3.3.2. Summary Statistics 119
3.3.3. Examiner Variation 128
3.4. Estimating Examiner Causal Effects 131
3.4.1. Research Design Overview 132
3.4.2. Computing Examiner Causal Effects 133
3.4.3. Results 136
3.4.4. Robustness Checks 140
3.5. Distinguishing Between Mechanisms 144
3.5.1. Predictive Regressions Methodology 144
3.5.2. Results 146
3.5.3. Robustness Checks 155

[2] Co-authored with Josh Feng.

3.6. Policy Implications 161
3.6.1. Calibrations 161
3.6.2. Discussion 165
3.7. Conclusion 167

4. The Lifecycle of Inventors[3] **169**

4.1. Introduction 169
4.2. Data 176
4.2.1. Patent Data 176
4.2.2. IRS Data 177
4.2.3. Matching 177
4.2.4. Inter-Generational Analysis 178
4.2.5. Test Score Analysis 178
4.2.6. Initial Data Description 179
4.3. Birth and Early Experience 179
4.3.1. Parental income 179
4.3.2. Gender 186
4.3.3. Race 188
4.3.4. Implications for Models of Talent Allocation 190
4.4. Inventors Pre- Labor Market 193
4.4.1. Schooling 193
4.4.2. Exposure to Innovation 195
4.5. Inventors in the Labor Market 205
4.5.1. Wages and Income 205
4.5.2. Inventor Age 209
4.5.3. Firms 210

[3]Co-authored with Alex Bell, Raj Chetty, Neviana Petkova and John Van Reenen.

4.5.4. Summary 210
4.6. A Simple Model of Inventors' Careers 211
4.6.1. Basic Model 211
4.6.2. Alternative Interpretations of Exposure Measures 213
4.6.3. Policies to improve entry into innovation 215
4.6.4. How much do tax policies on inventor's income affect innovation? . . 216
4.7. Conclusion 221

Bibliography **223**

A. Appendix to Chapter 1 **230**
A.1. More on Theory 230
A.2. More on Data 235
A.3. Estimation of Quality-Adjusted Inflation and Further Robustnes Checks . . . 237
A.4. Robustness Checks on the Relationship between Changes in Market Size and Quality-Adjusted Inflation 252

B. Appendix to Chapter 2 **261**
B.1. Additional Summary Satistics on Matched Inventors 261
B.2. Robustness Checks on the Causal Effect of Co-inventor's Premature Death . 265
B.3. Additional Results on Mechanism 277
B.4. Econometric Considerations 287
B.5. Data Appendix 289

C. Appendix to Chapter 3 **298**
C.1. More Descriptive Statistics 298
C.2. Random Assignment and Selection Effects 303
C.3. Bayesian and Frequentist Interpretations of Shrinkage Methodology 309

C.4. Additional Results on the Causal Effect of Patent Examiners 312
C.5. Additional Results on Mechanisms 323
C.6. Data Appendix 324

D. Appendix to Chapter 4 **326**
D.1. Appendix Figures and Tables 326
D.2. Data Appendix 340
D.3. Lifecycle Model of Innovation 344

List of Tables

1.1. Robustness of Inflation Difference Across the Income Distribution (Overlapping Products) For Various Periods and Departments 22

1.2. Decomposition of the Inflation Difference Between High- and Low-Income Households 24

1.3. Decomposition of the Inflation Difference Between High- and Low-Income Households Relative to Across-UPC Benchmark 26

1.4. New Products Target Higher-Income Consumers 31

1.5. Distribution of Estimated Module-Level Elasticites of Substitution For Three Income Groups 33

1.6. Differences in Price Level Paid for Same UPC by High- and Low-Income Households ($) 34

1.7. How Much of the the Difference in Inflation Between High and Poor is Explained by Patterns of Product Innovations? 46

1.8. Changes in Wholesale Costs vs. Changes in Retailer Margins 47

1.9. Causal Effects of Changes in Market Size 54

1.10. Robustness of Causal Effects of Changes in Market Size 55

2.1. Summary Statistics 73

2.2. Causal Effects of Co-inventor Death 83

2.3. Causal Effects of Inventor Death on Coworkers and Second-degree Connections 91
2.4. Distribution of Annual Changes in Log Total Earnings Before Co-inventor Death 93
2.5. Heterogeneity in Effect by Relative Ability Levels of Co-inventors 95
2.6. Collaboration Patterns Between Deceased and Survivor Inventors Before Death 96
2.7. Heterogeneity in Effect by Intensity of Collaboration Between Deceased and Survivor Inventors 98
2.8. The Causal Effect of Co-inventor Death On the Survivor Beyond Joint Production 100
2.9. The Causal Effect of Co-inventor Death across Firm and Geographic Boundaries 101

3.1. Art unit by year level statistics 120
3.2. Examiner level summary statistics 121
3.3. NPE Patent Holdings by NBER Technology Category 123
3.4. Aggregate NPE portfolio patent counts by initial assignee 124
3.5. Summary Statistics: NPE vs. Non-NPE Patents 126
3.6. NPE Examiner Effect Distribution Characteristics 138
3.7. Non-NPE Purchased Examiner Effects Distribution Characteristics 139
3.8. Robustness checks for the NPE examiner effect distribution. 141
3.9. Pairwise Examiner Effects Results - Purchase 149
3.10. Pairwise Examiner Effects Results - Litigation 152
3.11. Pairwise Differential Outcome Regressions 154
3.12. Robustness checks for key predictors NPE purchase. 157
3.13. Robustness checks for the effect of Examiner Career Effects on computed NPE effect distribution. 158
3.14. Horse Race Regressions Predicting Each Outcome 160
3.15. Simulating the Benefits of the Enhanced Patent Quality Initiative 164
3.16. Simulating the Benefits of the "Second-pair-of-eyes" Policy 165

4.1. Fraction of Gap in Patenting by Parental Income Accounted for by Differences in Math Test Scores 185
4.2. Fraction of Gap in Patenting by Gender Accounted for by Differences in Math Test Scores 187
4.3. Children's Patent Rates vs. Patent Rates in Father's Industry 199
4.4. Children's Patent Rates vs. Patent Rates in Neighborhood 202

A.1. Average Annual Inflation Rates Across Three Income Groups 241
A.2. Average Annual Inflation Rates Across Four Income Groups 242
A.3. Average Annual Inflation Rates across the Income Distribution at UPC*Geography Level 243
A.4. Average Annual Inflation Rates across the Income Distribution at UPC*Store Level 243
A.5. Average Annual Inflation Rates across the Income Distribution at Quarterly Level 243
A.6. Products that are about to exit have a lower inflation rate 244
A.7. Products that are about to exit have a higher price level 245
A.8. Share of spending on new and discontinued products across the income distribution 245
A.9. Decomposing the Difference in Sahares of Spending between High- and Low-Income Households 250
A.10. Isolating the Contribution of Stores and Local Market to the Overall Inflation Difference 252
A.11. Further Robustness Checks on Causal Effects of Changes in Market Size . . 255
A.12. Causal Effects of Changes in Market Size (Local Level) 257

B.1. Balance in Technology Classes For Survivor Co-inventors 262

B.2. Additional Balance Tests 263
B.3. Summary Statistics for Real and Placebo Coworkers and Second-Degree Connections 264
B.4. Balance for Number of Real and Placebo Survivor Coworkers per Deceased (Full Sample) 265
B.5. Testing For Dynamic Effects, P-Values of F-Tests 266
B.6. Balanced Panel of Survivors Experiencing Co-inventor Death between 2003 and 2008 266
B.7. Dynamic Causal Effect of Co-inventor Death, Full Sample 267
B.8. Dynamic Causal Effect of Co-inventor Death, Sample Restricted to Deaths from 2003 to 2005 268
B.9. Results for Main Outcomes, Excluding Deceased who Claimed Any Tax Deduction for High Medical Expenditures 270
B.10. The Causal Effect of Co-Inventor Death, Reweighting on the Propensity Score 272
B.11. Other Citation Metrics 273
B.12. Citation Results with Negative Binomial Estimator and Individual Fixed Effects 274
B.13. Testing For Differences Across Technology Classes 275
B.14. Inference on The Causal Effect of Co-Inventor Death Accounting For the Matching Step 276
B.15. Causal Effect of Coworker Death, Including Coworkers in Firms of Any Size 281
B.16. Sample Sizes for Analysis by Relative Ability Levels 281
B.17. Causal Effect of Co-Inventor Death on the Probability of Changing Firm . . 282
B.18. Causal Effect of Co-Inventor Death on the Probability of Getting a New Co-inventor 283
B.19. Heterogeneity in Causal Effect of Co-Inventor Death by Age Quartile 284
B.20. Heterogeneity in Causal Effect of Co-Inventor Death by Firm Size Quartile . 285

B.21. Heterogeneity in Causal Effect of Co-Inventor Death by Survivor's Citizenship Status 286
B.22. Heterogeneity in Causal Effect of Co-Inventor Death by Survivor's Network Size 287
B.23. Distribution of Outcomes for Two-Inventor Teams in 2002 (N=23,210) 296

C.1. Correlation of Blocking Actions 298
C.2. Art unit level statistics on the 670 art units in the PAIR data. Rates are unweighted, and are computed for units with more than 50 total cases in the period 300
C.3. Patents Asserted in Litigation (non-NPE) 301
C.4. NPE Patent Holdings by NBER Technology Category (RPX data) 302
C.5. Random Assignment Tests in Preferred Sample 303
C.6. Random Assignment Tests in Full Sample 304
C.7. Pairwise Examiner Alternate Blocking Action Effects Results 314
C.8. Pairwise Examiner Alternate Blocking Action Effects Results 315
C.9. Pairwise Examiner Effect Results for Other Legal Outcomes 316
C.10. Pairwise Examiner Blocking Action Effects Results (Application Level) 317
C.11. Instrumental Variables Results 321
C.12. Blocking Action Instrumental Variables Results 322
C.13. Results across technology centers for key predictors of NPE purchase and non-NPE litigation. 323

D.1. Additional Summary Statistics 326
D.2. Regression Estimates for Relationship between Math and Verbal Test Scores 327
D.3. Percent of Innovation Gap Accounted for by Test Scores - Alternative Estimators 328
D.4. Illustration of Technology Class and Distance 329

D.5. Patent Rates by Commuting Zone Where Child Grew Up 330
D.6. Percent of Female Inventors Across CZs 331

List of Figures

1.1. Inflation Across the Income Distribution (Overlapping Products) 19
1.2. Robustness of Inflation Difference between High- and Low-Income Households For Various Price Indices (Overlapping Products) 20
1.3. Inflation across Brand Price Deciles within Modules 25
1.4. Product Variety Increases Faster In Product Modules Catering to Higher-Income Households 29
1.5. Does the Price Fall When Demand Rises? 38
1.6. The Equilibrium Response of Price to Changes in Market Size Depends on the Price Elasticity of Demand 42
1.7. The Negative Relationship Between Inflation and Share of Spending on New Products 44
1.8. Changes in Wholesale Costs vs. Changes in Retailer Margins 48
1.9. Product Modules Catering to Higher Income Households Have Faster Growth and Lower Inflation 49
1.10. Changes in Market Size from Changes in the Age Distribution 51
1.11. Higher Market Size Leads to Product Entry 52
1.12. Higher Market Size Leads to Lower Inflation (Overlapping Goods) 53
1.13. The Inflation Difference Between High- and Low-Income Increases as Inequality Increases Faster 57

2.1. Team Composition for Two-Inventor Teams in 2002 74
2.2. Path of Outcomes around Co-inventor Death 77
2.3. Dynamic Causal Effects of Co-inventor Death 82
2.4. Path of Total Earnings for Survivors with Co-inventor Death in 2003-2005 . 85
2.5. Path of Outcomes for Coworkers and Second-Degree Connections Around Year of Death 90

3.1. Preliminary Evidence on the Relationship Between Examiner Tendencies and Patent Outcomes 129
3.2. NPE Examiner Effect Distribution 138
3.3. Distribution of NPE Examiner Effects (Empirical Bayes Count Model) . . . 143
3.4. Graphical representation of the mechanism results for pairwise prediction of patent purchases 150

4.1. Probability of Patenting by Age 30 vs Parent Income Percentile 180
4.2. Distribution of Math Test Scores in 3rd Grade for Children of Low vs. High Income Parents 182
4.3. Probability of Patenting by Age 30 vs 3rd Grade Math Test Scores 183
4.4. Percentage of Female Patent Holders by Birth Cohort 186
4.5. Distribution of Math Test Scores in 3rd Grade for Males vs. Females 188
4.6. Patent Rates by Race and Ethnicity in New York City Public Schools 189
4.7. Patent Rates vs. Test Scores by Race 190
4.8. Math Test Scores Conditional on Inventing 192
4.9. Probability of Patenting by Age 30 vs Parent Income Percentile in 10 Most Innovative Colleges 195
4.10. Technology Class-Level Patent Rates by Distance from Father's Technology Class for Children of Inventors 198

4.11. The Geographic Origin of Inventors 201
4.12. Patent Rates of Children Growing Up in a Commuting Zone vs. Patent Rates of Adults in that Commuting Zone 202
4.13. Percent of Inventors who are Female by State where Child Grew Up 204
4.14. Distribution of Inventors' Mean Individual Income Between Ages 40 and 50 . 205
4.15. Dynamics of Income Around Patent Application 207
4.16. Distribution of Inventors' Mean Individual Income Between Ages 40 and 50 . 208
4.17. Distribution of Inventors' Mean Individual Income Between Ages 40 and 50 . 219

A.1. Map of the 76 Local Markets Tracked in the Nielsen Datasets 236
A.2. The Mapping Between Nominal Income and Utility 237
A.3. Inflation Difference Between Various Income Groups For Various Price Indices (Fixed Basket) 244
A.4. The Distribution of Average Unit Prices Paid is the Same Across the Income Distribution (Reweighting by Spending Shares) 245
A.5. The Positive Relationship Between Share of Spending on New Products and Mean Consumer Income 246
A.6. The Relationship Between Share of Spending on New Products and Mean Income Depends on the Quality Engel Curves Elasticity 247
A.7. The Relationship Between Share of Spending on New Products and Representative Consumer Mean Income, Controlling for Household Fixed Effects . 248
A.8. Robustness Checks on Increase in Product Variety across the Income Distribution 249
A.9. Elasticities of Substitution Differ Across the Income Distribution 250
A.10. Quality-Adjusted Inflation Across the Income Distribution 251
A.11. The Share of Rich Households is Primarily Correlated with High Price Quantiles of New Products 252

A.12. The Share of Rich Households is Correlated with High Price Quantiles of New Products only in Product Modules with Steep Quality Engel Curves 253
A.13. Predicted and Actual Increase in Market Size 254
A.14. Causal Effects of Changes in Market Size (Local Level) 258
A.15. Geography Design: Predicting Market Size 259
A.16. Geography Design: Predicted Market Size Growth and Fixed-Basket Inflation 259
A.17. Do Product Innovations Follow Market Size or Change in Market Size? . . . 260

B.1. Number of Deceased Inventors Per Year 261
B.2. Tax Deductions for High Medical Expenditures Claimed by the Deceased . . 269
B.3. Path of Outcomes for Real and Placebo Survivor, Propensity Score Reweighting 271
B.4. Path of Outcomes for Coworkers and Second-Degree Connections around Year of Death 280
B.5. Comparison of Census Firm Size and Treasury EIN Size Distributions, 2002 295
B.6. Frequency of Collaborations Across EINs 297

C.1. Preliminary Evidence of Examiner Impact 299
C.2. Random Assignment Tests in Preferred Sample 305
C.3. Random Assignment Tests in Full Sample 306
C.4. Testing for Selection Effects 306
C.5. Distributions of examiner effects and tendencies by various outcomes observed in our dataset. 312
C.6. Comparison of Examiner Ranks 313
C.7. NPEs Tend to Purchase Patents from Examiners Who Are Reversed at PTAB 324

D.1. Patent Rates between Ages 30-40 vs. Parent Income Percentile 332
D.2. Patent Rates vs. Parent Income Percentile 333

D.3. Percentage of Children in Top 1% of Cohort's Income Distribution vs. Parent Income Percentile 334
D.4. Patent Rates vs. Parent Income in NYC Public Schools 335
D.5. Concentration of Innovation Among Graduates of Selected Colleges 336
D.6. Innovation at Age 30 Rates by College 337
D.7. Innovativeness of College vs. Parent Income Percentile 338
D.8. Patent Rate vs. Parent Income Percentile and Inventor Status 339
D.9. A Calibration of the Potential Gains from Supply-Side Policies (% Increase in Inventor Population) 357

Acknowledgments

Perhaps the most important lesson I learned from writing this dissertation is that research is a gradual, iterative and *collective* process. I was fortunate to benefit from the advice of many exceptional academics. First and foremost, I express my warm thanks to Raj Chetty and Philippe Aghion, Ed Glaeser, Larry Katz, Josh Lerner and Marc Melitz for their unwavering support and continuous guidance. I have learned a lot from their complementary perspectives on the economy and on economics. I hope to build on their insights for years to come, pursuing a research agenda at the intersection of macro questions and micro data. Moreover, I am indebted to Gary Chamberlain, Andrei Shleifer, Daron Acemoglu, Bill Kerr, Nathan Hendren and Sendhil Mullainathan for many conversations that shaped my thinking.

I would also like to thank my co-authors: the second chapter of this dissertation is co-authored with Neviana Petkova and Alex Bell, the third chapter with Josh Feng, and the fourth chapter with Alex Bell, Raj Chetty, Neviana Petkova and John van Reenen. I thank them for permission to include this joint work in my thesis.

The doctoral office at Harvard Business School and the Kauffman Foundation provided generous financial support as I was working on this dissertation. I am indebted to Jen Mucciarone, Karla Cohen, John Korn and Brenda Piquet for their help with all financial and logistical aspects of graduate school. Beyond Harvard, I thank Alan Marco for opening me the door to the United States Patent Office, and Bob Zider for introducing me to the world of innovation in Silicon Valley.

I am grateful to many friends in Cambridge who have made graduate school a very enriching experience, both academically and personally. I am especially thinking about Yuanjian Carla Li, Matt Basilico, Laura Blattner, Johann Blauth, Kirill Borusyak, John Coglianese, Elaine Chung, Greg Conti, Tilman Dette, Gabrielle Du Marais, Max Eber, Anastassia Fedyk and Giacomo Hodson, Josh Feng, Peter Ganong, Andy Garin, Sid George, Leander Heldring, Simon Jaeger, Sungho Kimlee, Danial Lashkari, Annie Liang, Jonathan Libgober, Luca Maini, Hannes Malmberg, Filippo Mezzanotti, Karl Neumar, Pascal Noel, Erik Nordbye, Daniel Pollman, Brendan Price, Laura Quinby, Alexandra Roulet, Alex Segura, Nihar Shah, Jann Spiess, Ludwig Straub, Tamer Teker, Edoardo Teso, Gabriel Unger, Jack Willis, Chenzi Xu and George Yin. I also thank Alexis Bonnardon, who in spite of the distance has always been a great influence on me.

Finally, I thank my father for instilling in me a thirst for knowledge, my mother for her unconditional support, and my grandparents for their constant encouragement. I dedicate this dissertation to the memory of my grandfather.

1. The Unequal Gains from Product Innovations: Evidence from the Retail Sector

1.1. Introduction

The rising level of nominal income inequality in the United States over the past thirty years has been a key topic of interest for academics and policymakers. The effect of skill-biased technical change in this process has been widely studied: innovations such as the diffusion of information and communications technologies have affected the relative price of skills and resulted in higher nominal income inequality. Much less attention has been paid to how price changes in the product market and the introduction of new products may differentially affect households at different points of the income distribution. Yet it is well-known that preferences are non-homothetic. Depending on their income level, households consume very different goods and services. Due to price changes in the product market over time, as well as changes in product variety, trends in real inequality may therefore differ from trends in nominal inequality. Product innovations may play a central role in this process by increasing the variety and quality of goods available in specific consumer segments, as well as by driving down the price of existing products in these segments due to increased competitive pressure. This paper shows the relevance of this hypothesis in the US retail sector over the past ten years, a large sector accounting for over 25% of the US economy.

I investigate this question in two steps. First, in the "measurement" part of the paper, I show that in the retail sector over the past ten years the quality-adjusted price index of high-income households rose substantially *slower* than that of low-income households, which amplified inequality. I build income-group-specific quality-adjusted price indices using detailed product-level data, in which I observe consumption patterns across income groups, price changes for all products available in consecutive years (inflation) and changes in product variety (product entry and exit). Second, in the "mechanism" part of the paper, I find that firms' equilibrium response to changes in demand across the income distribution explains why the price index of high-income consumers rose slower than that of the low-income. Specifically, my analysis shows that because demand from the high-income grew faster during this period, firms strategically introduced more new products catering to these consumers, which in turn drove down the price of existing products in these segments due to competitive dynamics.[1] The retail sector is ideal to conduct this investigation because it accounts for a

[1]A particularly good example illustrating this idea is the market for snacks. In recent years, meat snacks have grown tremendously, especially premium beef jerky - with sustained double-digit growth for over five years nationwide. Premium beef jerky is a high-protein, low-fat and low-calorie snack - a practical and healthy snack that particularly appeals to young and high-income households. The branding of premium beef jerky is fundamentally different from that of traditional jerky - favorite of truckers and staple of gas-station checkouts - and so is its production process. In particular, many of the varieties of premium beef jerky are fully organic - for instance, beef jerky made from 100% grass-fed cattle from networks of small family farms. The so-called "jerky renaissance" is largely driven by demand. It is answering the demand of high-income consumers concerned with healthy living and eager to support a sustainable, more humane agriculture. And it is taking place in a broader context of increased demand for snacks - a Nielsen survey found that one in ten Americans say they eat snacks instead of meals - and for proteins - according to the NPD group, more than half of Americans say they want more protein in their diet. The competition for the premium beef jerky market has intensified in recent years, with an ever increasing number of small, local players but also with the entry of established companies through acquisitions. For instance, Krave, one of the early players in premium jerky who led the market in the late 2000s, was acquired in 2015 by Hershey's, the largest chocolate manufacturer in North America. Accordingly, premium beef jerky prices have fallen and varieties have increased. Similar - although less spectacular - dynamics are visible in other segments of the snack industry, like hummus and protein bars, but not so in segments catering to lower-income consumers, like chips, bars and nuts.

large share of US GDP, rich data is available, and the notion of product (barcode) is well defined.

In the first part of the paper, I establish two new facts about inflation and increasing product variety across the household income distribution in the US retail sector from 2004 to 2013. I find that higher-income households experienced lower inflation and a faster increase in product variety than more modest households. The magnitude of these effects is large: over this period, the average annual inflation rate was 0.65 percentage point lower for households making more than \$100,000 a year, compared with households making less than \$30,000.[2] These results are very robust and hold before, during and after the Great Recession across product groups for a wide variety of price indices. They are based on detailed product-level data from the Nielsen Homescan Consumer Panel and Retail Scanner datasets, which are representative of the retail sector as a whole (which itself represents 40% of household expenditures on goods and 16% of household total expenditures). Whether similar results hold in other sectors of the economy is uncertain, but the analysis delivers a general methodological lesson for the measurement of inflation by statistical agencies: I show that the difference in inflation rates across the income distribution can be accurately measured only with product-level data. Indeed, a large share of the inflation difference between income groups occurs within detailed product categories, which cannot be captured by price series based on data aggregated at a level similar to what the Bureau of Labor Statistics and other statistical agencies currently use. These findings challenge the result from the existing literature that inflation is similar across the income distribution[3] and suggest that trends in real inequality may be diverging from trends in nominal inequality. Collecting product-level data is key to accurately measure this divergence. This has important potential policy implications given

[2]As discussed in Section 4, increasing product variety is valuable on its own, but empirically most of the welfare difference between households across the income distribution are captured by price changes in the basket of products that are available across years.

[3]See Section 4 for a detailed discussion of how my results relate to the literature.

the indexation of many government transfers.

In the second part of the paper, I examine whether the equilibrium response of supply to faster growth of demand from high-income consumers can explain the new facts on differential inflation and increase in product variety across the income distribution. It is a natural hypothesis to investigate because it is well known that in recent decades the share of national income accruing to high-income consumers has steadily increased (e.g. Song et al. 2016). To do so, I introduce a micro-founded model featuring monopolistic competition with variable elasticity of substitution preferences that differ across income groups, which generates a set of precise predictions that I take to the data and for which I find strong support. Intuitively, firms respond to changes in relative market size by skewing product introductions toward market segments that are growing faster. This process leads to a decrease in the price of existing products in the fast-growing market segments because increased competitive pressure from new products pushes markups down, and also because firms endogenously engage in process innovations lowering the marginal cost of these products. In my data, product groups catering to higher-income households grow faster and have a higher rate of product introduction, as well as lower inflation on existing products, which provides suggestive evidence in support of the theory but does not establish causality from demand to supply. To test the causal claim that increases in demand lead to a fall in inflation and an increase in product variety, I use shifts in the national age and income distribution between 2004 and 2013 to estimate the causal effect of changes in the number of consumers (market thickness) in a given part of the product space on inflation and product innovations. This research design is similar in spirit to Acemoglu and Linn (2004).

Taken together, my results show that in response to growing demand the equilibrium price of existing products falls and the rate of introduction of new products increases, and that these effects are sufficiently strong to explain the divergence in price indices across the income distribution. According to my point estimates, a 1 percentage point increase in demand leads to a 10 basis point decline in inflation and a 35 basis point increase in spending on new products. In line with the model, the magnitude of the effect is similar regardless of whether the change in demand comes from a change in the number of consumers or in per capita spending. In simple calibrations, I show that these effects are large enough to explain the new facts documented in the first part of the paper. In other words, these results suggest that absent the endogenous response of supply to market size effects, there would not have been a substantial difference in inflation nor in the rate of increase in product variety across the income distribution. This analysis has important implications for the endogenous growth literature, by providing evidence for endogenous product innovations across detailed product categories. It is also relevant for the trade literature and the debate on the role of markups in the gains from trade, because I establish empirically that the gains from increased market size are largely due to a fall in markups (consistent with the model and variable elasticity of substitution preferences). More broadly, these results are relevant for policy, given that the effectiveness of any government transfer crucially depends on the equilibrium response of supply to market size.

Overall, this paper provides new evidence challenging the existing literature primarily in two respects. First, the literature suggests that households across the income distribution tend to experience similar inflation rates (e.g. McGranahan and Paulson (2005)), except during peculiar periods like the Great Recession (Argente and Lee (2015)). Second, theoretical work has focused on the "product cycle", the idea that innovation is driven by economies of scale and allows for a trickle-down process bringing to the mass market the new products that were initially enjoyed by a select few at the top of the income distribution. In other words, innovation dynamics should contribute to lower quality-adjusted inflation for lower-income households. My findings suggest that market size effects may be a more important force, contributing to lower quality-adjusted inflation for higher-income households because market size grows faster at the top of the income distribution. More generally, this paper contributes to various strands of literature studying income inequality, price indices, technical change and monopolistic competition dynamics. [4]

The remainder of this paper is organized as follows: Section 1.2 describes the data, Section 1.3 presents the first main contribution of the paper, the measurement of quality-adjusted inflation and increasing product variety across the income distribution, and Section 1.4 makes the second main contribution of the paper, establishing that the endogneous response of supply to market size effects is very strong. A number of theoretical results, estimation details and robustness checks are reported in appendices.

1.2. Data Sources

1.2.1. Scanner Data

The analysis is primarily based on the Nielsen Homescan Consumer Panel and Nielsen Retail Scanner datasets, which have been widely used in the literature (Einav, Leibtag and Nevo, 2008). With this data, I can track consumption from 2004 to 2013 at the product level in department stores, grocery stores, drug stores, convenience stores and other similar retail outlets across the US. The data are representative of about 40% of household expenditures on goods and 16% of total household expenditures. Appendix B presents a detailed description of the data sources.

[4] More precisely, this paper relates to at least seven strands of literature, which respectively examine nominal income inequality (Autor, Katz and Kruger (1998), Autor, Katz and Kearney (2008), Piketty (2013), Song, Price, Guvenen, Bloom and Till von Wachter (2015), Atkinson (2015)), homothetic price indices (Sato (1976), Vartia (1976), Feenstra (1992), Pakes (2003), Broda and Weinstein (2006, 2010), Erickson and Pakes (2011), Comin, Lashkari and Mestieri (2015)), non-homothetic price indices (McGranahan and Paulson (2005), Broda and Romalis (2009), Moretti (2013), Diamond (2015), Handbury (2015), Faber and Fally (2015), Argente and Lee (2015)), innovation in labor markets (Acemoglu (1996, 2002, 2007), Acemoglu and Autor (2011), Autor (2013), Autor and Dorn (2013), Bell, Chetty, Jaravel, Petkova and Van Reenen (2015)), market size effects and endogenous technical change (Acemoglu and Linn (2004)), innovation and inequality in product markets (Schumpeter (1942), Vernon (1966), Matsuyama (2002)), and trade models of monopolistic competition with free entry (Melitz (2003), Melitz and Ottaviano (2008), Zhelobodko, Kokovin, Parenting and Thisse (2012)).

Three features of the data are particulalry useful for my analysis. First, product-level data is available on both prices and quantities. Quantity data is rare at the product level (for instance, the Bureau of Labor Statistics (BLS) does not collect such data) but it is crucial for quality adjustment in price indices. Intuitively, observing shifts in quantities allows to directly measure substitution patterns (and thus address substitution bias, which is a core concern of the CPI produced by BLS) and to infer the quality of products given their price, their market share, and the demand system. The quantity and price data is also used to structurally estimate the relevant parameters of the demand system in Section 1.3. Second, the Homescan Consumer panel has information on household characteristics such as income, age, education, size, occupation, marital status and zip code. It is therefore possible to directly map products to consumer characteristics. Third, the dataset offers a good measure of product innovations, defined as the introduction of new barcodes. Broda and Weinstein (2010) provide a detailed explanation regarding why it is reasonable to assume that all goods with different UPCs differ in some substantial way that might cause consumers to pay a different price for them and that it is rare for a meaningful quality change to occur that does not result in a change of UPC.[5] In other words, it is safe to assume that if the bar code changes, it is likely that some noticeable characteristic of the product has changed.[6] Similarly, it is possible to track products (barcodes) that are discontinued. Appendix Table A.8 shows that creation is larger than destruction, i.e. new products tend to steal market

[5]Broda and Weinstein (2010) make this point as follows: "Although it is difficult to enforce how a company uses a bar code, most industry experts strongly caution firms not to use the same bar code on more than one product. Doing so could cause confusion among retailers who would have trouble knowing what they were selling and for consumers whose receipts would not match their actual purchases. Similarly, firms typically do not use multiple UPCs for the same product because that makes it very difficult for retailers to reorder out of stock items. As a result, manufacturers tend to use other bar code systems for internal use and reserve the UPC for tracking products that are identical to the consumer. For example, changing the slogan on a Heinz ketchup bottle does not require a new bar code, but changing the size of the bottle does."

[6]Note that these measures of product turnover include any change in products, including those driven by changes in the size of products, their flavor, or other characteristics that can be secondary for the consumer. Nielsen provides identifiers that allows tracking barcodes that are new just because of a change in size or flavor: all of the results presented in the paper are similar when excluding these products from the definition of "new" products.

shares from existing products. As will become apparent in Section 1.3, this displacement is indicative of biases in conventional price indexes that ignore the effects of changing quality.

Nielsen provides a detailed product hierarchy, based on where products are sold in stores. In my sample, about 3 million products (identified by their barcode, or "UPC") are classified into 10 broad departments (dry grocery, general merchandise, health and beauty care, alcoholic beverages, deli, ...), 125 more detailed product groups (grooming aids, soup, beer, pet care, kitchen gadgets, ...) and 1,075 very detailed product modules (ricotta cheese, pet litter liners, bathroom scale, tomato puree, women's hair coloring, ...). When ranking product modules by mean consumer income, the top five product modules are scotch, natural cheese, gin, fondue sauce and cookware, while the bottom five are tobacco, canned meat, taco filing, insecticide and frozen fruit drinks.

Finally, the data can be disaggregated at the level of 76 local markets, described in Appendix B. According to Nielsen, the dataset is still representative within each of the 76 markets. The data cannot reliably be disaggregated further (e.g. at the county or zip code level).

1.2.2. Markup Data

I also have access to weekly product-level data between January 2004 and June 2007 in 19 U.S. states, for 250 grocery stores operated by a single retail chain. This dataset contains information for 125,048 unique products (UPCs), mostly in the food and beverages categories, housekeeping supplies, books and magazines, and personal care products. Most of the stores are located in the western and eastern corridors, in the Chicago area, Colorado and Texas. For every store in every week, data is available on the price, the wholesale cost and the marginal cost of each product. I infer the markups of the retailer based on the information on the price and wholesale cost. Note that I do not measure other costs like labor, rent and utilities. In the analysis carried out in Section 1.4, store-year fixed effects are used to absorb these costs. The dataset also reports "adjusted gross profits" per unit for each product, defined as the net price minus the sum of wholesale costs and transportation costs plus net rebates from the manufacturer - I use this adjustment in robustness checks.

1.2.3. Manufacturer Data

Finally, I have obtained data from GS1, the company in charge of allocating bar codes in the US, on the universe of barcodes and manufacturers. I match the bar codes observed in the Nielsen data to manufacturers using the first few digits of the bar code - the match rate is close to 95%. Since the cutoff size for a manufacturer to appear in this dataset is to make a sale rather than an arbitrary number of workers, I can observe the full distribution of manufacturers in each product group. There are about 500 manufacturers on average in each product group, with 90 percent of the product groups having more than 200 manufacturers. The median number of products supplied by a manufacturers is 5 and the average is 14.

Consistent with the findings reported by Hottman et al. (2016), while on average half of all output in a product group is produced by just five manufacturers, around 98 percent of manufacturers have market shares below 2 percent. Thus, the typical product group is characterized by a few large manufacturers and a competitive fringe of manufacturers with very low market shares. A second important feature of the data is that even the largest manufacturers are not close to being monopolists: the largest manufacturers in a product group on average has a market share of 22 percent. The model presented in Section 1.4 is consistent with these patterns.

1.3. Measuring Quality-Adjusted Inflation Across the Income Distribution

In this section, I compute quality-adjusted inflation rates across the income distribution, taking into account the welfare gains from increasing product variety. I start with a brief reminder on non-homothetic preferences and price indices. Second, I follow standard results in the literature to derive the exact price index in my preferred demand system. Third, I present results and robustness checks for inflation on existing goods across the income distribution. I show the relevance of these results for statistical agencies like BLS and discuss differences with the existing literature. Fourth, I document the difference in changes in product variety (due to both product creation and destruction) across the income distribution. Finally, I bring together the findings on inflation on overlapping products and on product creation and destruction to compute the full quality-adjusted inflation rate.

1.3.1. Nonhomothetic Preferences, Product Variety and Real Inequality

The nonhomothetic nature of preferences means that the baskets of goods and services consumed by households across the income distribution systematically differ. Given that households have a taste for variety, the mapping between nominal income and utility depends on both the quality-adjusted price of products and the number of available varieties. This paper studies how the mapping between nominal income and inequality changes over time. Figure A.2 illustrates this idea. In this example, the "new" mapping is an upward shift of the "old" mapping (for instance because of productivity gains), but the shift is asymmetric and benefits higher-income households relatively more. The shift takes into account changes in the quality-adjusted price of products as well as changes in the variety of available products for each nominal income level.

This paper characterizes shifts in the mapping from nominal income to utility at various points of the income distribution using a money metric, the compensating variation. The compensating variation gives the amount of nominal income that one would need to take away from the consumer at the "new" equilibrium to make them indifferent between this new equilibrium (with the new mapping) and the "old" one (with the initial mapping). This approach provides a characterization of changes in real inequality. Given the demand system, it is possible to infer the quality of products based on their price and equilibrium market share, and to measure the gains from increasing product variety based on the share of spending on new products. The rest of this section discusses the procedure in detail and shows that the results are robust across price indices, indicating that structural assumptions about the demand system do not drive the results.

I use the term "inflation" to describe my findings throughout the paper because it is an intuitive notion, but my results are invariant to the unit of account. I document changes in the relative prices of goods that cater to high- and low-income households. These relative price changes would be unaffected by shifts in the overall level of inflation, therefore nominal indeterminacy plays no role in my findings.[7]

[7] It is also useful to note that given that the set of goods is not fixed, the difference in the rates of quality-adjusted inflation experienced between high- and low-income households could be permanent. If the set of goods were fixed, the divergence in inflation rates between goods should be bounded and eventually converge to 0, otherwise in the long run all consumers, regardless of their income level, would switch to the goods with slower price increases. But since there is entry and exit, quality-adjusted inflation may be permanently lower for one income group relative to another (e.g. at any point in time the price of the products catering to the high income may remain higher than that of the products catering to the low income, but in a quality-adjusted sense the price of the high-end products may be very low).

1.3.2. Overview of Methodology and Review of Basic Price Indices

The goal is to compute the cost of achieving a certain level of utility in one year relative the previous year. Such price indices are known as “exact price indices.” This requires taking into account changes in product quality, product variety, as well as the optimizing behavior of consumers who may substitute from one good to another. By definition, this exercise requires taking a stance on a utility function. The role of the utility function is twofold: quantifying the impact on utility of price changes for the goods that exist across periods, but also translating into a welfare metric the patterns of product creation and destruction. In order to understand what parts of the result are driven by structural assumptions on the utility function, it is useful to split this analysis into two parts, first considering price changes on products that exist across periods and second considering changes in product variety.

First, I consider inflation on the set of products available in two consecutive years. The quality of a given product is assumed to be constant over time[8] and data is available on market shares of each product, therefore it is straightforward to compute a price index reflecting product quality and consumers' substitution behavior. Intuitively, I observe the price change for each product and I only need to decide how to weight the various products. The exact price index offers a principled way of doing so. The structural assumption on the utility function play a minor role for the final result, as can be seen by computing standard price indices that do not have an interpretation in terms of utility but can serve as bounds by allowing for an extreme form of substitution (like the Paasche price index, which offers a lower bound on inflation) or making any substitution impossible (like the Laspeyres price index, which offers an upper bound on inflation). In addition to the exact price index derived

[8]This assumption is standard in the literature: see for instance Feenstra (1994) and Broda and Weinstein (2010). It appears to be reasonable even though advertising or the introduction of complementary goods may violate it. It can be tested, and I present the results of this test in the robustness test section below.

in the following subsection, I consider the following price indices :

$$\begin{aligned}
Laspeyres\ Index: \quad P_L &\equiv \frac{\sum_{i=1}^n p_i^t q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} = \sum_{i=1}^n \frac{p_i^t}{p_i^0} s_i^0 \\
Paasche\ Index: \quad P_P &\equiv \frac{\sum_{i=1}^n p_i^t q_i^t}{\sum_{i=1}^n p_i^0 q_i^t} = \left(\sum_{i=1}^n \left(\frac{p_i^t}{p_i^0} \right)^{-1} s_i^t \right)^{-1} \\
Marshall - Edgeworth\ Index: \quad P_{ME} &\equiv \frac{\sum_{i=1}^n p_i^t (q_i^t + q_i^0)}{\sum_{i=1}^n p_i^0 (q_i^t + q_i^0)} \\
Walsh\ Index: \quad P_W &\equiv \frac{\sum_{i=1}^n p_i^t \sqrt{q_i^t q_i^0}}{\sum_{i=1}^n p_i^0 \sqrt{q_i^t q_i^0}} \\
Fisher\ Index: \quad P_F &\equiv \sqrt{P_L P_P} \\
Geometric\ Laspeyres\ Index: \quad P_L^G &\equiv \Pi_{i=1}^n \left(\frac{p_i^t}{p_i^0} \right)^{s_i^0} = exp \left(\sum_{i=1}^n s_i^0 \cdot log \left(\frac{p_i^t}{p_i^0} \right) \right) \\
Geometric\ Paasche\ Index: \quad P_P^G &\equiv \Pi_{i=1}^n \left(\frac{p_i^t}{p_i^0} \right)^{s_i^t} = exp \left(\sum_{i=1}^n s_i^t \cdot log \left(\frac{p_i^t}{p_i^0} \right) \right) \\
Tornqvist\ Index: \quad P_T &\equiv \Pi_{i=1}^n \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^t + s_i^0}{2}} = exp \left(\sum_{i=1}^n \frac{s_i^0 + s_i^t}{2} \cdot log \left(\frac{p_i^t}{p_i^0} \right) \right)
\end{aligned}$$

Another relevant consideration is whether or not to chain the price index. In a chain index, each link consists of an index in which each period is compared with the preceding one, the weight and price reference being moved forward each period. A chain index is therefore path dependent: it depends on the prices and quantities in all the intervening periods between the first and last period in the index series. When there is a gradual economic transition from the first to the last period, chaining is advantageous because it smoothes trends in relative prices and quantities and tends to reduce the index number spread between the various price indices listed above. But if there are fluctuations in the prices and quantities in the intervening periods, chaining may not only increase the index number spread but also distort the measure of the overall change between the first and last periods.[9] Accordingly, I

[9]For example, suppose all the prices in the last period return to their initial levels in period 0, which implies

present robustness check with and without chaining the indices.

Second, I follow standard techniques in the literature to provide an adjustment to the price index depending on the rate of increase in product variety. For new product and discontinued products, price relatives are not available. Intuitively, given that consumers have a taste for variety, an increase in the range of available product should lead to a decrease in the price index. Translating the increase in product variety into welfare gains requires structural assumptions. The key assumption is that within a product module, varieties are horizontally differentiated. I use two standard frameworks: nested CES utility (see next subsection) and nested translog utility (in progress). Importantly, these structural assumptions do not matter quantitatively because the elasticity of substitution of products within a module is very high, such that the gains from increasing product variety are already reflected in the prices of existing products (this is shown formally in the next subsection). The key point is that the patterns of product creation and destruction matter in general equilibrium, but their welfare effect is almost entirely taken into account in the price changes of products existing across periods. I also provide bounds showing that the patterns of product creation and destruction in the data will lower the price index more for high-income households than low-income households and, even under some violations of the structural assumptions.

Although the main text focuses on results and robustness checks across income groups (various price indices, various subsamples of products and years, etc.), I have also conducted this analysis across age groups. The results of this analysis are reported in Appendix. If find that inflation tends to be higher for older individuals, which is in line with the market size hypothesis examined in Section 1.4.

that they must have flucutated in between. A chain Laspeyres index will not return to 100: it will tend to be greater than 100. If the cycle is repeated with all the prices periodically returning to their original levels, a chain Laspeyres index will tend to drift further and further above 100 even though there may be no long-term upward trend in the prices. Chaining is therefore not advised when the price fluctuates.

1.3.3. Estimation Framework for Nested CES Exact Price Index

The estimation framework builds on Feenstra (1994) and Broda and Weinstein (2006, 2010). I split the analysis using three representative agents, one for households making less than \$30,000 a year, one for households making between \$30,000 and \$100,000 a year, and one for households making above \$100,000. Preference parameters in my estimation framework are a flexible function of the income level, which allows for nonhomotheticicities.

The remainder of this subsection shows how to derive and estimate the price index for any representative agent. I assume a nested CES utility function, following Feenstra . Product groups are indexed by g and G is the set of all product groups. The elasticity of susbtitution across product groups is ρ. The elasticity of susbtitution between product groups is $\sigma = \rho/(\rho-1)$. The upper level utility function is:

$U = (\sum_{g \in G}(C_{gt})^{\rho})^{\frac{1}{\rho}}$

Composite consumption within a product group is given by:

$C_{gt} = (\sum_{m \in M_g}(c_{mgt})^{\rho_g})^{\frac{1}{\rho_g}}$

where $\sigma_g = \rho_g/(\rho_g - 1)$ is the elasticity of substitution between product modules within product group g.

$c_{mgt} = (\sum_{u \in U_m}(d_{umgt}c_{umgt})^{\rho_m})^{\frac{1}{\rho_m}}$

where c_{ubgt} is the consumption of UPC u in product module m and product group g in period t. $\sigma_m = \rho_m/(\rho_m - 1)$ between UPCs within product module m. d_{umgt} is unobserved and reflects the quality of the UPC. So we want to estimate σ and two high-dimensional sets of elasticities of substitution, $\{\sigma_g\}_g$ and $\{\sigma_m\}_m$. We expect $\sigma_m > \sigma_g$ since there is more substitution across UPCs within a module than across modules within a group.

The minumum unit cost function of the subutility function at the product module level is:

$P_{mgt} = (\sum_{u \in U_{mgt}}(\frac{p_{umgt}}{d_{umgt}})^{\sigma_m})^{\frac{1}{\sigma_m}}$

The minimum cost function at the product group level is:

$P_{gt} = (\sum_{m \in M_g} (P_{mgt})^{\sigma_g})^{\frac{1}{\sigma_g}}$

And the overall price index is given by:

$P_t = [\sum_g P_{gt}^{\sigma}]^{\frac{1}{\sigma}}$

Consumer optimization also yields:

$s_{umgt} = \left(\frac{p_{umgt}/d_{umgt}}{P_{mgt}}\right)^{1-\sigma_m}$

i.e. the quality adjusted price can be backed out as follows:

$ln\frac{p_{umgt}}{d_{umgt}} = \frac{ln(s_{umgt})}{1-\sigma_m} + ln(P_{mgt})$

The key insight for estimation is that the share of consumption of UPC u depends directly on the quality-adjusted price. We can write the price index only in terms of prices and market shares even when goods are constantly being replaced.

If we make the assumption that product quality is constant over time ($d_{umgt} = d_{umgt-1}$) and if ignore the introduction of new products, given our assumption of a (nested) CES utility function and the results in Sato (1976) and Vartia (1976), the exact price index is:

$P_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) = \Pi_{u \in I_{mg}} \left(\frac{p_{umgt}}{p_{umgt-1}}\right)^{w_{umgt}}$

where $I_{mg} = I_{mgt} \bigcap I_{mgt-1}$ is the set of varieties consumed in both periods t and $t-1$. x_{mgt} and x_{mgt-1} are the cost-minimizing quantity vectors of products within module m in each of the two periods. A remarkable feature is that the price index does not depend on the unknown quality parameters d_{umgt}. We only need to compute the geometric mean of the individual variety price changes, where the weights are ideal log-change weights. These weights are computed using cost shares in the two periods and are always bounded between the shares of spending in the t and $t-1$ (in other words the price index is bounded between the geometric Paasche and Laspeyres indices described in the previous subsection):

$s_{umgt} = \frac{p_{umgt} x_{umgt}}{\sum_{u \in I_{mg}} p_{umgt} x_{umgt}}$

$w_{umgt} = \frac{(s_{umgt} - s_{umgt-1})/(ln(s_{umgt}) - ln(s_{umgt-1}))}{\sum_{c \in I_{mg}} (s_{umgt} - s_{umgt-1})/(ln(s_{umgt}) - ln(s_{umgt-1}))}$

As shown in Broda and Weinstein (2010), with change in varieties across periods the exact price index (quality-adjusted inflation) for product module m within product group g is then given by:

$$\pi_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) = P_{mg}(p_{mgt}, p_{mgt-1}, x_{mgt}, x_{mgt-1}, I_{mg}) \cdot \left(\frac{\lambda_{mgt}}{\lambda_{mgt-1}}\right)^{\frac{1}{\sigma_m - 1}} \quad (1)$$

with

$$\lambda_{mgt} = \frac{\sum_{u \in I_{mg}} p_{umgt} x_{umgt}}{\sum_{u \in I_{mgt}} p_{umgt} x_{umgt}} \; ; \; \lambda_{umgt-1} = \frac{\sum_{u \in I_{mg}} p_{umgt-1} x_{umgt-1}}{\sum_{u \in I_{mgt-1}} p_{umgt-1} x_{umgt-1}}$$

This result states that the exact price index with variety change is equal to the "conventional" price index multiplied by an additional term, which captures the role of new and disappearing varieties. The higher the expenditure share of new varieties, the lower is λ_{mgt} and the smaller is the exact price index relative to the conventional price index. An intuitive way to rewrite this ratio is as follows:

$$\frac{\lambda_{mgt}}{\lambda_{mgt-1}} = \frac{1 + Growth\ Rate\ of\ Spending\ on\ Overlapping\ Products_{gmt}}{1 + Growth\ Rate\ of\ Total\ Spending_{gmt}}$$

which clearly shows that a net increase in product variety (weighted by spending) drives the price index down. The price index also depends on the module-specific elasticity of substitution between varieties σ_m. As σ_m grows, the additional term converges to one and the bias goes to zero: intuitively, when existing varieties are close substitutes to new or disappearing varieties, price changes in the set of existing products already take into account the entry of more varieties.[10]

[10] One can better understand the implications of the choice of time horizon by considering an examples of how the proposition captures the impact of different types of creation and destruction, quoting from Broda and Weinstein (2010):

"Let's consider the case of a new type of sunscreen that replaces an earlier type. If the new sunscreen is just a repackaging of last year's sunscreen without a noticeably different quality or price, then, ceteris paribus, the new sunscreen will have a market share equal to that of the old sunscreen. If this is true, then the share of common goods will be unchanged and our measured quality bias from the replacement of the old model would be zero. If, instead, the new sunscreen is priced identically but is of a higher quality than the old model, then, ceteris paribus, its market share will rise. This result comes directly from the optimizing behavior of the consumer, because the new sunscreen will have a lower price per unit quality than the old

In principle, we could use the result above to compute price indices adjusted for increasing product variety over any time horizon. However, two factors make some time horizons more sensible than others in practice. First, it makes sense to define periods in years to prevent seasonal factors from driving product turnover. Thus, UPCs will be considered destroyed only if they were not purchased at any time during a yearlong period. Second, we need to decide how many years should separate the two periods. While this choice is inherently arbitrary, I decided to present results based on one-year intervals, considering other intervals in robustness checks. As mentioned earlier, a key assumption is that the taste or quality parameters for common goods must remain constant in start and end years of the sample. In fact, it may vary over short horizons due to anything that might affect demand (e.g., marketing or fashion considerations). The reason for why immutable preferences over long time horizons must be assumed when deriving price indexes is that if the utility function is changing over time for either exogenous reasons (e.g., fashion) or endogenous reasons (e.g., marketing) then one cannot make sensible statements about how price changes affect welfare, nor can one derive exact price indexes because identical price vectors will yield different utility levels at different times. The choice of the time horizon also matters for the magnitude of the adjustment term for increasing product variety.

Thus, we need data on quantity and price for new products, discontinued products, and products existing across periods, which is readily available in the Nielsen data. We also need to estimate the two high-dimensional sets of elasticities of substitution, $\{\sigma_g\}_g$ and $\{\sigma_m\}_m$. The main challenge for estimation of is that we want to obtain a demand and supply equation using only information on prices and quantities. The insight of Feenstra (1994) as extended by Broda and Weinstein (2006) is that although we cannot identify supply and demand, the data does tell us something about the joint distribution of supply and demand parameters.

sunscreen. If this is the case, the higher share of the new good relative to the old good implies that there is a "quality bias" in the conventional price index that only considers products existing across periods."

Appendix C gives details about how to derive the estimation equations.[11]

1.3.4. Inflation Across the Income Distribution For Goods Available in Consecutive Years

1.3.4.1. Results

Figure 1.1.: Inflation Across the Income Distribution (Overlapping Products)

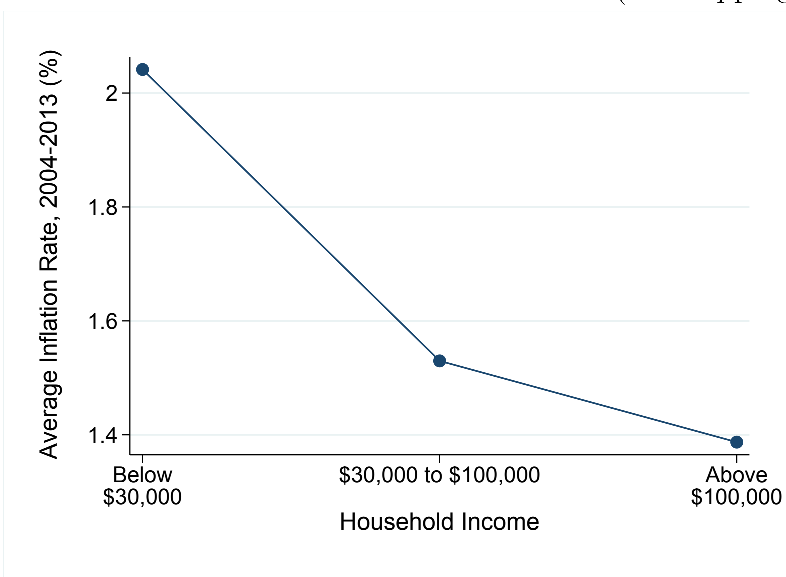


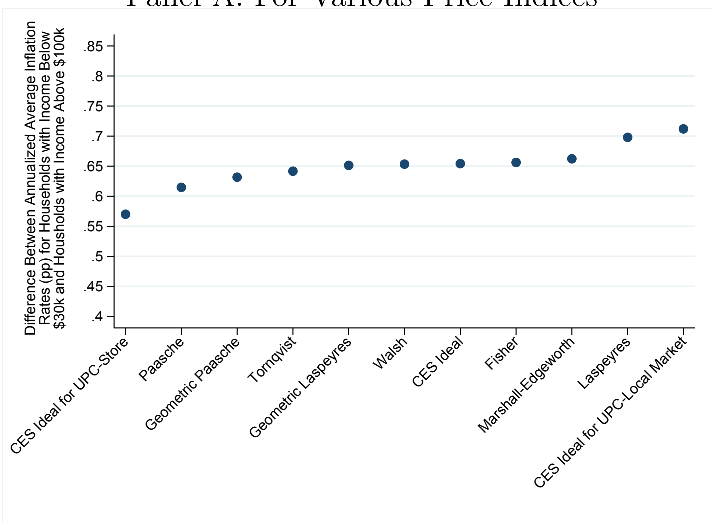
Figure 1.1 shows the average inflation between 2004 and 2013 on the set of overlapping products (defined as products that are available in consecutive years) for households across the income distribution. Inflation is computed using the exact price index for the nested CES utility function described in the previous subsection (without the adjustment for new and disappearing products, which is examined later in this section and does not affect the

[11] I have also estimated a demand system based on the translog expenditure function, following Feenstra and Weinstein (2015). The results are qualitatively similar to those presented here and are available from the author upon request.

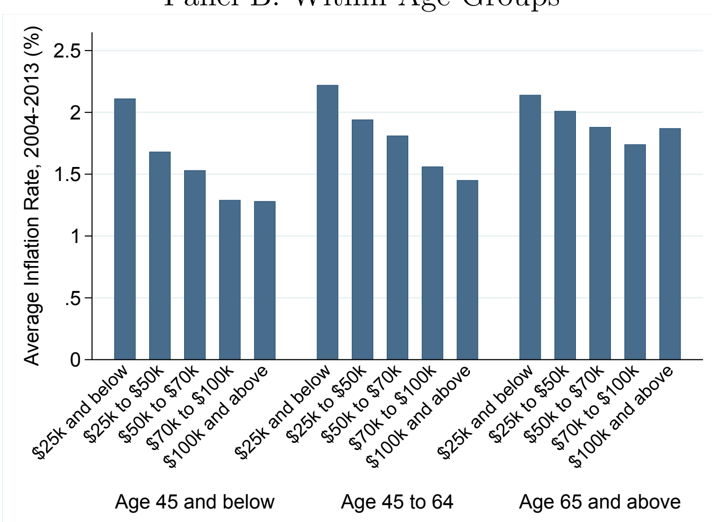
results). The inflation rate is about 0.65pp lower for households making more than $100,000 a year, relative to households making less than $30,000.

Figure 1.2.: Robustness of Inflation Difference between High- and Low-Income Households For Various Price Indices (Overlapping Products)

Panel A: For Various Price Indices



Panel B: Within Age Groups



As shown in Panel A of Figure 1.2, similar results are obtained when considering any of the price indices introduced in Subsection 1.3.2. In addition, Panel A of Figure 1.2 reports the inflation difference when re-defining products as UPCs available in the same store, or as UPCs available in the same local market (see Appendix B for a map of local markets). The results with this new definition of products are very similar. Overall, across all price indices and product definitions, the inflation rate is always between 0.56pp and 0.72pp lower for households making more than $100,000 a year, relative to households making less than $30,000. Panel B of Figure 1.2 shows that these results are robust when considering other income groups and when repeating the analysis within age groups. For each age group, inflation is sytematically lower for the higher-income households.[12]

Table 1.1 shows the robustness of this result across subsamples. The difference between the inflation rates of high- and low-income households exists before, during and after the Great Recession,[13] and it is not driven by any single department. Appendix C presents various additional robustness tables and figures. First, Tables A.1 and A.2 describe the level of inflation for various cuts of the income distribution, various price indices and various periods. Figure A.3 summarizes this information and shows that the difference in inflation rates is very robust: higher-income households consistently experienced a lower inflation rate. Second, I redefine products to be UPCs available in a given local market (Table A.3) or UPCs available in a given store (Table A.4) and show that the results continue to hold. Additional robustness checks are discussed at the end of this section.

[12] In a companion paper, Jaravel (2016) investigates patterns of inflation and product innovations across the age distribution.

[13] The difference inflation rates appears to be larger during the Great Recession. Argente and Lee (2015) argue that the way in which consumers adjusted their shopping behavior to mitigate the crisis can explain the difference in the inflation rates across the income distribution during this period.

Table 1.1.: Robustness of Inflation Difference Across the Income Distribution (Overlapping Products) For Various Periods and Departments

Period	Excluded Department	Average Annual Inflation Difference between High- and Low-Income Households
2004-2013	None	0.654
2004-2006	None	0.472
2011-2013	None	0.529
2004-2013	Health and beauty care	0.689
2004-2013	Dry grocery	0.738
2004-2013	Frozen food	0.690
2004-2013	Dairy	0.649
2004-2013	Deli	0.657
2004-2013	Packaged meat	0.654
2004-2013	Fresh produce	0.655
2004-2013	Non-food grocery	0.534
2004-2013	Alcohol	0.638
2004-2013	General merchandise	0.631

1.3.4.2. Decompositions

It is possible to decompose the inflation difference between households at difference points of the income distribution. For the purpose of this exercise, I focus on comparing households making more than $100,00 a year to households making less than $30,000 a year. The inflation difference reflects the combined effects of both price and quantity changes, as well as baseline differences in spending patterns across income groups. For instance, it could be that high-income households spend more on fresh produce and that inflation tends to be lower in this broad item category. Alternatively, it could be the case that high-income households experience different inflation rates compared with low-income households on the same bar codes, for instance because they shop at different stores or have different propensities to use coupons. Accordingly, the inflation difference between high income and low-income households can be decomposed into a “between” component, which corresponds to the inflation difference that would prevail if households differed only in terms of their expenditure shares and experienced the same inflation rate within an item category, and a "within" component, which corresponds to the inflation difference that would prevail if households differed only in terms of the inflation rate they experience within an item category and had the same expenditure shares across categories. Formally, for any grouping of products G, we can decompose the inflation difference between high- and low-incomme households as follows:

$$\pi^R - \pi^P \equiv \sum_G s_G^R \pi_G^R - \sum s_G^P \pi_G^P = \underbrace{\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G \right)}_{Between} + \underbrace{\sum_G \overline{s_G} (\pi_G^R - \pi_G^P)}_{Within}$$

with s_m^i the share of spending of income group i on product grouping G andπ_G^i the inflation experienced by income group i in product grouping G. π_G and $\overline{s_G}$ denote the average inflation rate and the average spending shares for product grouping G.

Table 1.2 reports the results of the decomposition at the following levels of aggregation: department, product group, product module, UPC, UPC in a given local market, and UPC in a given store. Inflation is directly observed at the product level for the last three categories, and the definitions of inflation for categories at levels of aggregation above the UPC are given in subsection 3.2. Perhaps not surprisingly, less than 10% of the differenced in the inflation rate experience by high- and low-income households is due to differences in spending across broad departments. More surprisingly, less than 25% of the inflation difference results from different spending patterns across the 125 detailed product groups, and less than 45% of the difference from spending patterns across the 1,025 very disaggregated product modules. More than 70% of the inflation difference occurs between UPCs. This is a large share of the overall difference in inflation rates, but a substantial fraction of the difference still occurs *within* UPCs. To assess the mechanism at play, I repeat the decomposition at the level of UPCs in a given local market, which brings the share of the "between" component close to 80%, as well as at the level of UPCs in a given store, which brings the share of the "between" component to 92%. Taken together, these results show that most of the difference in inflation rates between high- and low-income households occurs across UPCs, and that some of the effect results from differential price dynamics for the same UPC across stores. In Section 1.4, I examine whether local competition and changes in markups can explain these patterns.

Table 1.2.: Decomposition of the Inflation Difference Between High- and Low-Income Households

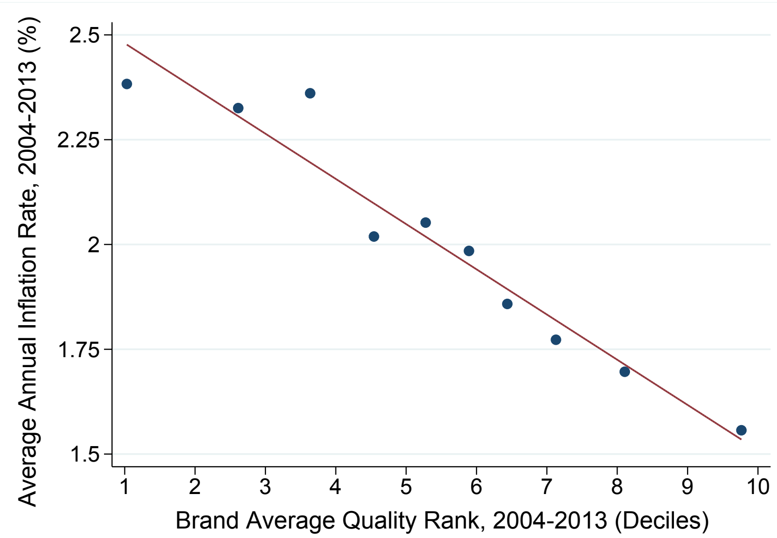
Aggregation Level (Broad to Narrow)	Decomposition	Inflation Difference pp	% of actual
Department	Between	0.06	8.6
Product Group	Between	0.14	21.4
Product Module	Between	0.28	42.8
UPC	Between	0.476	72.2
UPC-Local Market	Between	0.520	78.8
UPC-Store	Between	0.607	92.1

Table 1.3 delves further into the differences in price dynamics across UPCs, which are of particular interest because they explain most of the inflation difference across the income distribution and because the Nielsen data is less reliable to document variation in prices paid by different income groups for the same UPC. Indeed, Nielsen often automatically enters the price of the UPC based on the store the panelist reported for their shopping trip.[14] Figure 1.3 documents that within product modules inflation is lower for products that belong to brands in higher price deciles. The price deciles are computed within each module based on the average (spending-weighted) unit price of the products that belong to a brand, over all years in the dataset. This approach provides a way to segment the product space even within product modules, the highest level of disaggregation provided by Nielsen. It is not subject

[14]Nielsen obtains the price data from that store in that particular week.

to mean reversion because the deciles are not based on the price of the UPC itself, but rather on pricing behavior at the brand level over the entire dataset. Table 1.3 shows that differences in the spending patterns of high- and low-income households across price deciles within product modules explain more than 85% of the inflation difference between high and low-income households that exists across UPCs. Taken together, the decomposition exercises show that the inflation difference between high- and low- income households primarily exists across UPCs, rather than within, and that it can be accounted for by the fact that inflation is lower for products belonging to higher-quality brands, which primarily cater to higher-income consumers.

Figure 1.3.: Inflation across Brand Price Deciles within Modules



Laspeyres Inflation Rate. Coeff. -0.107*** (s.e. 0.008).

Table 1.3.: Decomposition of the Inflation Difference Between High- and Low-Income Households Relative to Across-UPC Benchmark

Aggregation Level (Broad to Narrow)	Decomposition	Inflation Difference pp	Inflation Difference % of benchmark
Department	Between	0.061	**12.8**
Product Group	Between	0.143	**30.0**
Product Module	Between	0.282	**59.2**
Product Module*Price Decile	Between	0.408	**85.7**
UPC	Between	0.476	**100**

1.3.4.3. Relevance for the Methodology of Statistical Agencies

Table 1.2 means that product-level data is needed to capture the magnitude of the difference in inflation rates between households at different points of the income distribution. It is not sufficient to simply reweight aggregate price series based on income-specific spending shares, even when the level of aggregation is as detailed as product modules. Yet this is precisely the approach followed by the BLS and other statistical agencies. More specifically, the BLS collects prices on 305 different item categories, known as "entry-level items" (ELI). Most of these item categories are very coarse. 230 of them are actually in the retail sector, where the level of disaggregation is much higher than in other sectors. Still, this level of aggregation is too high to capture the bulk of the difference between high and low income consumers. This explains why the result presented here may appear inconsistent with the existing literature, which has found small differences between high and low income consumers.

For instance, McGranahan and Paulson (2005) compute income-specific inflation rate based on between-ELI inflation differences and income-specific CEX spending patterns. Using their data, I computed that between 2004 and 2013 the annualized inflation difference for households in the bottom vs. top income quartiles was 0.18pp, which is similar to what I obtained in the Nielsen data with the "between product group" methodology (see Appendix C for details).

Therefore, the conventional wisdom that inflation is similar across the income distribution may be misplaced: statistical agencies like BLS collect data at a broad level of aggregation, which biases the estimate of the difference in inflation across income groups towards zero. Using the Nielsen data, I directly show that the magnitude of this bias is large in the retail sector. Table 1.3 suggests that a large share of the inflation difference across income groups could be captured by segmenting each of the detailed item categories by price deciles - the data collected by statistical agencies like the BLS could be used to replicate this approach.[15]

1.3.4.4. Related Literature

My results are consistent with Argente and Lee (2015), who study the inflation difference for high- and low-income households during the Great Recession, find that it is lower for higher-income households and argue that this effect is driven by substitution patterns. The inflation dynamics I describe in this paper are more general and of a different nature: I show that the difference in inflation rates across the income distribution extends well beyond the crisis and continues to hold even when substitution effects are ignored (indeed, Figure 1.2 shows that the magnitude of the inflation difference is similar across a variety of price indices that do not allow for substitution, like the Laspeyres index). In Section 1.4, I show that the magnitude of the inflation difference between high- and low-income households can be explained by the equilibrium response of supply to market size effects.[16]

[15] One would then need to infer the spending shares of various income groups along price deciles, which could be done for instance by estimating "quality Engel curves" as in Bils and Klenow (2001).

[16] Note that both my results and the results of Argente and Lee (2015) appear inconsistent with the findings of Broda and Romalis (2009), who also use Nielsen data and report in an unpublished manuscript that they find that inflation is lower for lower-income households.

Two other recent papers are closely related to my findings. Pisano and Stella (2015) document that lower-income households pay lower prices than higher-income households for the same products, primarily because they shop more at discount stores. In contrast, I focus on changes in income-specific price indices over time and use the demand system to provide a measure of quality-adjusted inflation. Faber and Fally (2015) explore the implications of firm heterogeneity for household price indices across the income distribution. They find that larger, more productive firms endogenously sort into catering to the taste of wealthier households, and that this gives rise to asymmetric effects on household price indices in their structural model. I provide direct evidence of differences in inflation rates across the income distribution and, in Section 1.4, I focus on a completely different explanatory mechanism.

To the best knowledge, my paper is the first to propose decompositions of the inflation differential between high- and low-income households as in Section 4.3.2. and to relate these patterns to the dynamics of product creation and endogenous changes in markups, which are discussed in the remainder of the paper. My analysis shows that collecting product-level data is key to accurately measure the divergence of inflation rates across the income distribution - this methodological lesson is likely to apply to other sectors beyond retail. The sign and magnitude of the inflation difference between high- and low-income households in other sectors remains an open question.[17]

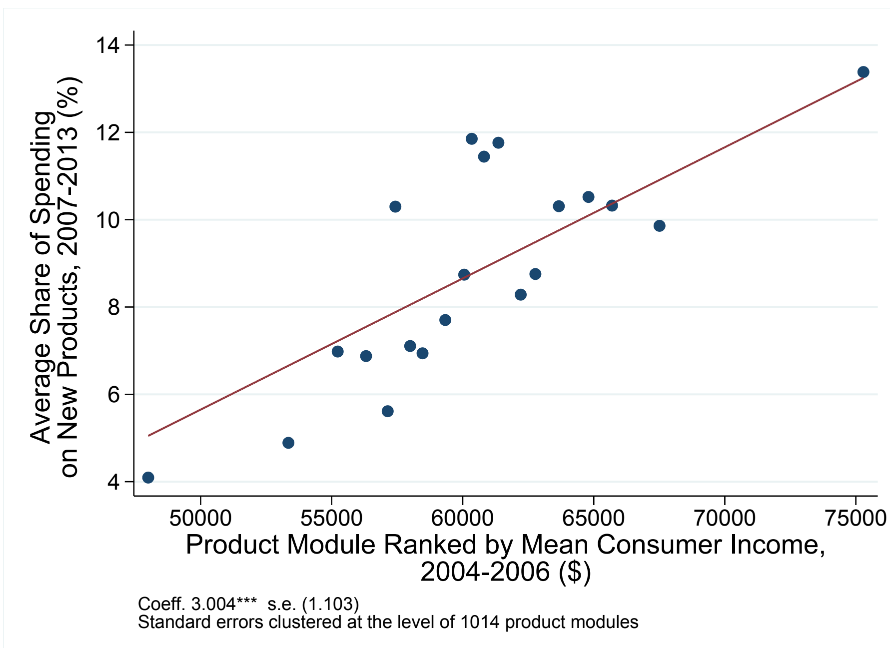
1.3.5. Changes in Product Variety Across the Income Distribution

Do welfare effects from increasing/decreasing product varieties also differ across the income distribution? I find that the rate of increase in product variety is faster in product modules

[17] See the debate between Moretti (2013) and Diamond (2016) about price changes in the housing sector for college and high-school graduates. To my knowledge, there have been very few attempts at developing non-homothetic price indices and McGranahan and Paulson (2005) remains the main reference. Handbury (2013) developed a non-homothetic price indices across cities but did not have panel data to study inflation dynamics.

catering to higher-income households, implying that higher-income households benefit more from increasing product variety. Figure 1.4 shows this effect by using the share of spending on new products (barcodes which did not exist in the previous year) as a measure of the flow of successful product innovations. For every $10,000 increase in the mean income of the consumers buying from a product module, the share of spending in this product module goes up by 3 percentage points, about a third of the average share of spending on new products. Appendix D reports patterns of product destruction across product modules, which are much more homogeneous.

Figure 1.4.: Product Variety Increases Faster In Product Modules Catering to Higher-Income Households



According to the demand system presented in Section 1.3, the share of spending on new products is the welfare-relevant metric. However, is supply or demand driving the relationship shown on Figure 1.4? It could be the case that more products are introduced in product modules catering to high-income consumers because of supply effects, which may be exogenous (e.g. it may be inherently easier to introduce new product at the high-end of the product space) or endogenous (e.g. if innovators and suppliers decide to specifically target higher-income consumers). Alternatively, it could be the case that higher-income consumers have a higher taste for novelty and purchase new products wherever they are introduced in the product space. In other words, the share of spending on new products may be higher in product modules catering to higher-income households simply because new products diffuse faster (while the rate of product introduction may be similar across modules). To isolate the contribution of supply, the ideal regression would compare the same household moving across the product space. Such a regression can be directly run in the Nielsen data, at the household H $\times$ product module M level with household fixed effects:

$$ShareSpendingNewProducts_{HM} = \alpha + \beta ProductModuleIncomeRank_M + \alpha_H + \epsilon_{HM}$$

where α_H is a household fixed effect and $ProductModuleIncomeRank_M$ is the rank of the product module by income of the representative consumer in the product module (computed using 2004-2006 data). The results are reported in Table 1.4, with standard errors clustered at the household level. As in the previous graphs, I find a strong positive relationship between the share of spending on new products and the mean income of the consumer in the product module. This analysis confirms that supply plays a role in this process, because household fixed effects ensure that the relationship is not driven by a composition effect across modules (i.e. different propensities of consumers to buy new products wherever they show up in the product space). I also present specifications with interaction terms for whether the household is "high-income" (income above $100,000) or "low-income" (income below $30,000). The magnitude of the interaction effects is small, around 10% of the effect for middle-income households.

Table 1.4.: New Products Target Higher-Income Consumers

	$ShareSpendingNewProducts_{hm}$	
$ProductModuleIncomeRank_M$	2.79***	2.82***
	(1.024)	(1.031)
$ProductModuleIncomeRank_M \times HighIncome_H$		-0.24***
		(0.063)
$ProductModuleIncomeRank_M \times LowIncome_H$		0.11*
		(0.058)
$Household\ Fixed\ Effects$	Yes	Yes

Standard errors clustered by product modules.

I have checked that similar results hold for other measures of "new products" - new UPCs relative to two, three or four years ago, as well as new brands. Appendix D reports additional results. In particular, Figure A.8 shows that, across product modules, the rate of increase in the total number of varieties increases by one percentage point with a $10,000 dollar increase in the income of the representative consumer. I have also investigated patterns of product destruction, which are much more homogeneous across the product space. Moreover, I have examined patterns of product creation and destruction within product modules, which are similar to the patterns observed across modules (i.e. even within a module, higher income households tend to benefit more from product creation and destruction). Table A.9 shows that the differences in shares of spending on new products between high- and low-income consumers largely occur within product modules - this pattern is very similar to the inflation decomposition discussed earlier and provides preliminary evidence that there is a tight connection between the inflation and product innovation patterns, which is further examined in Section 1.4.

1.3.6. Quality-Adjusted Inflation Across the Income Distribution

Using the results from equation (1), I can bring together the previous facts on inflation for products available in consecutive years and patterns of product creation and destruction. Table 1.5 shows the distribution of the estimated elasticities of substitution by income groups. Two findings stand out. First, the elasticities tend to be slightly smaller for higher-income households, i.e. higher-income households are less price elastic in equilibrium.[18] This provides direct evidence of non-homotheticities (in addition, Figure A.9 in Appendix C shows that the elasticites for high- and low-income consumers are not very strongly correlated). Second, the magnitude of the elasticities is very high. Using the optimal markup formula derived in Section 2, these magnitudes are consistent with the observed markups in the retail sector.[19] The high values of the elasticities means that the “product variety” adjustment is very small: since the elasticities of substitution are very high, most of the welfare effects are captured by the inflation difference on goods that exist across consecutive years (cf. derivation in subsection 4.2). As a result, the quality-adjusted inflation across the income distribution (Figure A.10 in Appendix C) looks virtually identical to inflation across the income distribution for overlapping products (Figure 1.1).

[18] Note that the equilibrium elasticity of susbtitution depends on consumers' preference parameters, but also on the competitive environment if the elasticity of substitution is not constant. See Section 3 for a discussion of models of monopolistic competition with variable elasticities of substitution. I have estimated a demand system based on the translog expenditure function, which features decreasing elasticities of substitution, and have obtained qualitatively similar results.

[19] In retail (groceries and food) the margin is around 2.71%.
See http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/margin.html

Table 1.5.: Distribution of Estimated Module-Level Elasticites of Substitution For Three Income Groups

		Income < \$30k	Income ∈ [\$30k-\$100k]	Income > \$100k
Percentile	10th	9.08	9.30	8.56
	25th	13.91	13.44	13.47
	50th	21.45	19.87	19.78
	75th	35.30	32.74	34.25
	90th	65.98	56.75	71.27

1.3.7. Further Robustness Checks

Selection effects. A potential concern is that the inflation patterns described above could result from selection effects, for instance if low-income households overwhelmingly consume goods whose characteristics are rendered obsolete by the entry of new products. In such a case, a relatively higher share of the goods consumed by the poor would be exiting the market in any given year - the price changes for these goods are not observed, but if they were they would be negative because these products face tougher competition.[20] Tables A.6 to A.8 show that such selection effects are in fact not at play in the data.[21]

Price Convergence Anoter potential concern is that the observed inflation difference between high- and low-income households could be driven by the fact that high-income households might initially pay a higher price for the same UPCs as low-income households, and the price would then converge to the same level for all households in future periods. The last three rows of Table 1.2 reject the hypothesis by showing that the "within-UPC" share of the total inflation difference is modest. A more direct way of showing that this mechanism is not the driving force, without the need for any assumption about the demand system, is to run a regression of the unit price of the UPC on a UPC fixed effect and an indicator

[20]See Pakes and Erickson (2011) for a discussion of such selection effects.

[21]Note that even if selection effects were at play, the nested CES structural demand system with new goods addresses these concerns by adjusting the price index when new varieties enter.

for whether the household is high income (restricting attention to products purchased by both income groups). Table 1.6 reports the results of such a regression and show that, in any given year, households making more than \$100,000 a year tend to pay about 2.9% more for the same UPC, compared with households making less than \$30,000 a year. This result is consistent with the findings of Pisano and Stella (2015). The magnitude of this effect is negligible compared with the 0.65pp difference in inflation rates, which over the course of a few years leads to a much bigger welfare difference between high- and low-income households than the difference in price levels in any given year.[22] Figure A.4 in Appendix C provides complementary evidence by showing that the distribution of average unit prices paid by high- and low-income households is very similar, restricting attention to the set of products purchased by both income groups.

Table 1.6.: Differences in Price Level Paid for Same UPC by High- and Low-Income Households (\$)

	Average Unit Price
High-Income Household	0.0664***
	(0.00118)
Constant	2.2825***
	(0.00061)
UPC*Year Fixed Effect	Yes
R^2	0.9954

The product cycle. One may worry that the patterns about inflation and new products are driven by the "product cycle" - namely, products start in the market with a very high price, and at that point are only purchased by high-income households, and then converge to their long-run, stable price, at which point they start being purchased by lower-income households. I address this concern in several ways. First, my results hold across the product space, as

[22] Note that my focus on inflation allows me to take into account changes in product variety and consumer substitution across products over time, as well as to characterize how these patterns differ across the income distribution. The static analysis of the levels of prices paid for the same barcodes by individuals across the income distribution does not speak to these dynamic consideration, which are first order in the data.

shown in Figures 1.4 and 1.9. If the product cycle was driving the results, then the measured differences in inflation and product innovation should only be visible from the point of view of each individual consumer and not across the product space. Second, I have repeated the analysis by considering only products in the middle of their lifecycle. Specifically, in any given year I have resticted the sample to products that had been in the market for at least two years and that would remain in the market for at least two more years. The inflation patterns obtained with this approach are similar to those reported above. Third, I have shown that the product cycle is not an important force in the data as barcodes do not travel down the income distribution (empirically, barcodes tend to remain in the same price decile during their entire lifecycle, which is intuitive for the retail sector and stands in contrast with other products like computers). Fourth, even if the product cycle was an important force in the data, under the assumption described at the beginning of this section the nested CES demand system will provide an accurate estimate of the quality-adjusted inflation rate for each of the various income groups, given the speed of the product cycle. In particular, in this analysis the "novelty" of a product is determined separately for each income group based on the basket of goods consumed by this income group in the previous year.

The fashion cycle. A distinct concern is that the inflation patterns may be driven by a phenomenon analogous to the "fashion cycle" - the fact that products exhibit seasonality patterns and that the price of older products falls disproportionately. For instance, because of the fashion cycle measured inflation is negative in the apparel industry - yet productivity gains for apparel are small and it would be incorrect to infer large welfare gains from the observed price patterns.[23] Conceptually, the fashion cycle means that the assumption that the "quality" of a barcode is fixed over time fails - if newness is a key feature of the utility derived from a product, the observed price of this product will fall over time but this may not reflect any change in the quality-adjusted price. I address the concern that high-income

[23]The Bureau of Labor Statistics addresses this by making hedonic adjustments and by ignoring sale prices.

households may be more like to be affected by the fashion cycle in two ways. First, the fashion cycle is about churn of products and not about a net increase in the number of available varieties. I show that there is a faster increase in varieties in the parts of the product space that cater to higher income households, but there is not more churn. Similarly, the price patterns across product modules are predicted by the net increase in product variety, rather than by churn. Second, the results hold even with product categories where the fashion cycle is unlikely to exist, such as food product .

Local market effects and store effects. Using a reweighting procedure, I show that the differences in spending patterns across stores and local markets cannot explain the inflation patterns previously documented. The results are reported in Appendix Table A.10.

Alternative measures of household income. I repeated the analysis with three alternative measures of household income: reported income divided by household size; total retail expenditures per capita within a household; and whether the head of household is a college graduate.

Sampling variability. To ensure that the results are not driven by differing degrees of sampling variability across income groups, I built a random subsample of the data with an equal number of households in each of the income bins. I have also checked that the results across product modules hold in the Retail Scanner Data (which is based on information recorded directly at the store, not obtained from households, and contains many more observations as described in Appendix B).

Base drift. I have repeated this analysis using unchained price indices instead of chained indices and obtained similar results.

Quarterly data. Table A.5 shows that the results are very similar when repeating the analysis at a quarterly frequency.

1.4. The Equilibrium Response of Supply to Changes in Demand

In this section, I investigate the hypothesis that an important driver of the results presented in Section 1.3 - namely, the fact that higher-income household experience lower inflation and a faster increase in product variety from 2004 to 2013 - is differential income growth and directed product innovations. I first present a theory showing how rising nominal income inequality may have a causal effect on the direction of product innovations, and in turn result in a further increase in real income inequality. An increase in market size leads to more product entry, which puts downward pressure on the prices of existing products through increased competition. I test the key channels of this theory, first by studying a series of descriptive patterns on price and markup dynamics and then by tracing out the causal impact of a demand shock on price and innovation dynamics. I then discuss other possible mechanisms.

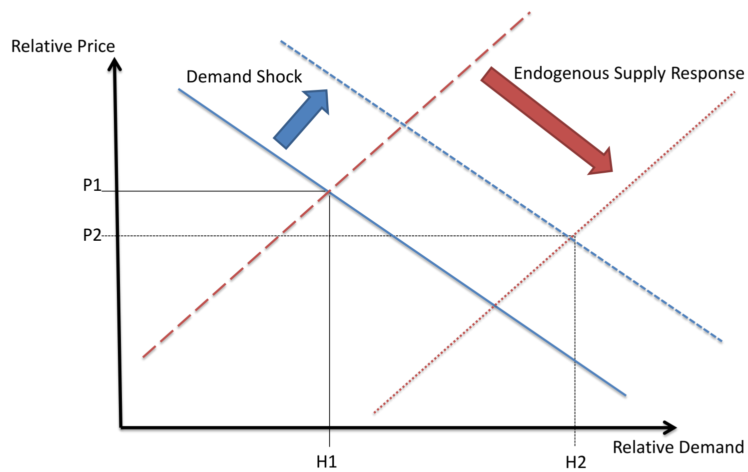
1.4.1. Theory

1.4.1.1. Intuition: Tracing Out the Observed Long-Term Supply Curve

Because of nonhomothetic preferences and the endogenous price changes induced by changes in relative demand, changes in nominal inequality may overstates or understate changes in real inequality. Consider Figure 1.5. When relative demand goes up, if the short-run supply curve is upward sloping as in standard price theory, then the equilibrium price should go up. However, supply may endogenously shift out due to the response of firms to market size effects, which will at least mitigate the price increase and, as illustrated in Figure 1.5, could potentially result in a new equilibrium price that is lower than the initial equilibrium price. This "price overshooting" case is shown to be the empirically-relevant case in Section 1.4.

In other words, the observed long-term supply curve is downward sloping.[24]

Figure 1.5.: Does the Price Fall When Demand Rises?



To investigate whether changes in nominal inequality overstate or understate changes in real inequality, the following concepts are useful:

- **Weak equilibrium (relative) bias** ("directed technical change"): when demand for a good becomes relatively more abundant, supply (technology, innovation, entrepreneurship, etc.) becomes endogenously biased towards this factor.
- **Strong equilibrium (relative) bias**: the relative supply curves for goods are downward sloping.

Consider demand H for a high-quality good and demand L for a low-quality good. Endogenous technology A is a function of relative demand $\frac{H}{L}$. The equilibrium relative price is $\frac{p_H}{p_L} = f\left(\frac{H}{L}, A(\frac{H}{L})\right)$.

[24]The observed long-term supply curve is defined as the nexus of equilibrium points traced out by shifts in the demand curve.

There is weak equilibrium bias if:

$\frac{\partial f}{\partial A}\frac{\partial A}{\partial H} < 0$

There is strong equilibrium bias if:

$\frac{\partial f}{\partial H} + \frac{\partial f}{\partial A}\frac{\partial A}{\partial H} < 0$

where $\frac{\partial f}{\partial H} > 0$, as in standard price theory.

The equations above and 1.5 provide an intuitive reduced-form way of thinking about the effect of shifts in demand on the equilibrium price. In the next subsection, I discuss a specific microfounded model that is consistent with the evidence, generates additional predictions and provides a framework for estimation and for welfare calculations.

1.4.1.2. A Microfoundation

I focus on microfounded models of monopolistic competition with free entry of products. This broad class of models is appealing for three reasons: the assumption of monopolistic competition is reasonable in retail (the Herfindahl index for most product groups is below 0.20), these models nest the standard model of directed technical change (Acemoglu, 2002), and they generate rich product-level predictions and counterfactuals. The intuition for the effect of changes in market size on supply in monopolistic competition models is as follows: an increase in market size leads to more product entry, which puts downward pressure on the prices of existing products (pecuniary externality). Therefore, in such models innovation occurs entirely through product entry - there is no “process innovation” reducing the marginal cost of the existing products, whose price dynamics are determined by changes in markups.

Within the class of monopolistic competition models with free entry of products, only some models are consistent with the “price overshooting” case illustrated in Figure 1.5. In particular, the CES model of Acemoglu (2002) does not allow for the possibility that the price goes down when demand goes up (see Appendix A for a detailed derivation). On the other hand, Melitz and Ottaviano (2008) is consistent with the strong equilibrium bias (see Appendix A for a derivation). In the rest of this section, I characterize the conditions under which "price overshooting" is possible using the general monopolistic competition model of Zhelobodko, Kokovin, Parenting and Thisse (2012). The key insight is that, in general equilibrium, the curvature of the utility function and variable markups drive the sign and magnitude of the response of the equilibrium price to changes in market size.

L consumers with additively separable preferences over varieties solve:

$max_{x_i \geq 0}\, U = \int_0^N u(x_i)di \qquad s.t. \qquad \int_0^N p_i x_i di = E$

Consumer maximization yields

$p_i(x_i) = \frac{u'(x_i)}{\lambda}$

$\lambda = \frac{\int_0^N x_i u'(x_i)di}{E}$

Total quantity demanded is $q_i = Lx_i$. The monopolist takes the residual demand curve as given and solves:

$max\ \pi(q_i) = R(q_i) - C(q_i) \equiv \frac{u'(q_i/L)}{\lambda}q_i - V(q_i) - F$

with $V(.)$ is the variable cost function and F the fixed cost. The optimal markup of the producer is therefore given by:

$M^* = -\frac{x_i \cdot u''(x_i)}{u'(x_i)}$

At the free entry equilibrium, $\pi(q_i^*) = 0$ and a mass N^* of firms satisfies labor market clearing:[25]

$N^* = \frac{L \cdot E}{C(q_i^*)}$

[25] A similar model can be solved by assuming that the sector is small relative to the total economy, which allows for ignoring some GE effects. See Mayer, Melitz and Ottaviano (2016).

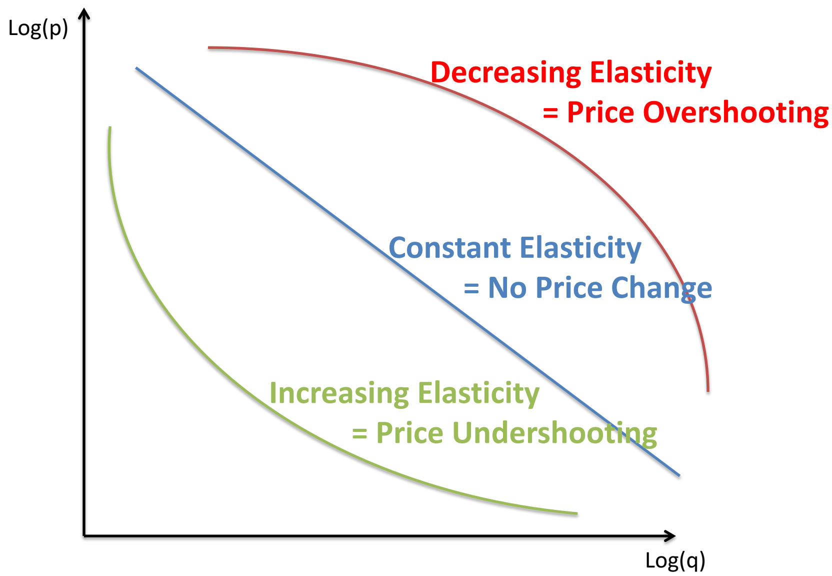
Therefore, the model delivers the following comparative statics:

$\frac{dN^*}{dL} > 0 \qquad \frac{dx_i^*}{dL} < 0 \qquad \frac{dM_i^*}{dL} \lesseqqgtr 0$

Therefore, the optimal markup is given by the inverse of the price elasticity of demand[26] (which, given our assumption of separability, is equal to the inverse of the elasticity of substitution between varieties). This result is very general and holds regardless of the shape of the cost function $V(.)$. It shows why the equilibrium response of prices to changes in market size crucially depends on the curvature of the utility function. The intuition for the comparative statics is as follows. When market size increases, new products enter the market. As a result, consumers start spreading out their expenditures across more products, due to taste for variety. Consequently, consumption per capita x_i for the existing products goes down, which induces a responses of the optimal markup M^*. The equilibrium markup which may increase, decrease or stay unchanged, depending on the properties of demand. Figure 1.6 shows this effect in log-log space. The blue curve corresponds to CES demand, as in Acemoglu (2002). Movements along the curve do not matter, the elasticity is constant. On ther other hand, the red curve shows that when consumption per capita decreases (moving to the left along the curve), the price elasticity of demand goes up, i.e. the optimal markup goes down. Melitz and Ottaviano (2008) corresponds to this case. Conversely, as shown with the green curve, if the price elasticity of demand is increasing the equilibrium price should go up in response to an increase in market size.

[26] This term is also equal to the coefficient of relative risk aversion.

Figure 1.6.: The Equilibrium Response of Price to Changes in Market Size Depends on the Price Elasticity of Demand



Because preferences are nonhomothetic, the curvature of the utility function may differ for consumers in different income groups and the equilibrium response of price to market size may be different in product modules catering to different consumer segments. This framework allows for rich counterfactuals to answer the question: what would have been the difference in inflation and rates of product introduction across the income distribution absent the endogenous response of supply to market size effects? The framework is based on homothetic utility functions within product modules, but I can separately estimate these utility functions for different groups of consumers across the income distribution, which effectively allows for nonhomotheticities by letting the parameters of the utility function vary freely with the level of income.[27]

[27]Section 1.3 shows how to nest the various sub-utility functions for each product module into one aggregate

In sum, the model makes three predictions: 1. there is a strong negative correlation between inflation and the share of spending on new products both *across* and *within* product modules; 2. the inflation patterns across the income distribution are driven by differences in changes in markups; 3. growing demand causes more product innovations and lower inflation. I test and find support for all three predictions in the rest of this section.[28]

1.4.2. The Relationship Between Inflation and Share of Spending on New Products

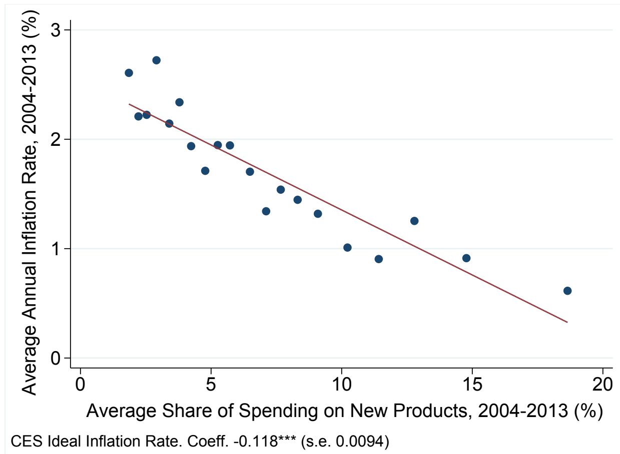
The negative correlation between inflation and the share of spending on new products is a key feature of the data. Panel A of Figure 1.7 shows this relationship across product modules. Panel B shows that the relationship persists within product modules when segmenting the product space by price deciles: product entry is higher in higher price deciles, while inflation is lower.

utility function.

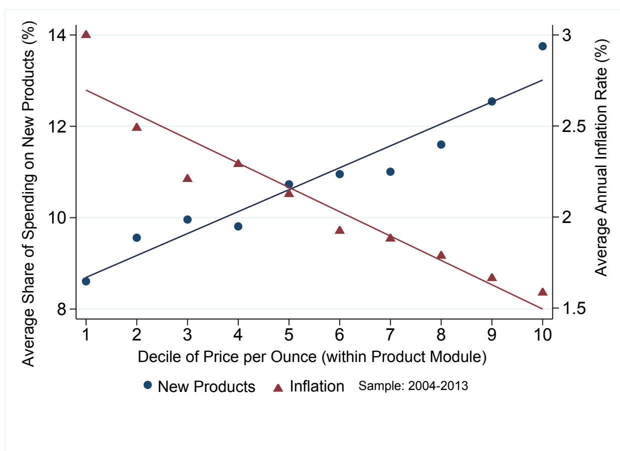
[28]Note that these tests speak to an active debate in the trade literature about the source of the gains from trade and the role of variable markups and variable elasticity of substitution preferences. See in particular DeLoecker, Goldberg, Pavcnik and Khandelwal (2012), Feenstra and Weinstein (2016), and Mayer, Melitz and Ottaviano (2016)

Figure 1.7.: The Negative Relationship Between Inflation and Share of Spending on New Products

Panel A: Across Product Modules



Panel B: Within Product Modules



A simple decomposition exercise shows that the relationship between inflation and product innovations across modules can explain a large fraction of the inflation patterns across income groups documented in Section 1.3.[29] As previously mentioned, for any product grouping G, we can write the inflation difference between income groups as:

[29] This is similar in spirit to the reweighting technique introduced in DiNardo, Fortin and Lemieux (1996).

$$\pi^R - \pi^P \equiv \sum_G s_G^R \pi_G^R - \sum s_G^P \pi_G^P = \underbrace{\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G\right)}_{Between} + \underbrace{\sum_G \overline{s_G}(\pi_G^R - \pi_G^P)}_{Within}$$

with s_m^i the share of spending of income group i on product grouping G,π_G^i the inflation experienced by income group i in product grouping G, and with π_G and $\overline{s_G}$ denoting the average inflation rate and the average spending shares for product grouping G. We can now decompose the "between" component further and examine how much of the inflation difference across categories is explained by (or predicted by) differences in shares of spending on new products across categories:

$$\left(\sum_G s_G^R \pi_G - \sum s_G^P \pi_G\right) = \left(\hat{\pi_G^R} - \hat{\pi_G^P}\right) + R$$

with

$$\hat{\pi_G^R} - \hat{\pi_G^P} = \sum_G \hat{\beta} X_G (s_G^R - s_G^P)$$

$$R = \sum_G \hat{\epsilon_G}(s_G^R - s_G^P)$$

$$\pi_G = \beta X_G + \epsilon_G$$

where X_G is share of spending on new products in G. $\hat{\beta}$ is the OLS estimate of β. This procedure calibrates the extent to which the difference in inflation rates between high- and low-income households results from the fact that high-income consumers tend to devote a higher share of their spending to product categories where the rate of product innovations is higher (i.e. moving to the right along the x-axis in panel A of Figure 1.7), or from the fact that high-income households tend to spend more on product categories with a lower share of inflation, holding the rate of product innovations constant (i.e. moving down the y-axis in panel B of Figure 1.7). Table 1.7 shows that for the various levels of aggregation, around half of the inflation difference between high- and low-income households can be explained by differences in patterns of product innovations.[30] These results provide strong support for the

[30] Note that any measurement error (e.g. UPC relabeling that does not reflect a true product innovation) will bias this esitmate downward, therefore these estimates can be viewed as a lower bound.

first prediction of the model and the notion that the joint dynamics of product innovation and inflation are crucial to understand changes in real inequality.

Table 1.7.: How Much of the the Difference in Inflation Between High and Poor is Explained by Patterns of Product Innovations?

Aggregation Level (Broad to Narrow)	Share of Rich-Poor Inflation Difference Explained
Department	40.9
Product Group	58.3
Product Module	51.3

The relationships described so far are only correlations and should not be interpreted as causal, but they provide transparent evidence on the pervasive nature of the relationship between inflation and product innovations and on its relevance for understanding changes in real inequality.

1.4.3. The Role of Markups

As mentioned in Section 1.2, I observe retailer price p_{it} and wholesale cost c_{it} from 2004 to 2007 for a subset of the product. A first-order Taylor expansion yields a convenient additive expression for the log price change:

$p_{it} = m_{it} + c_{it}$

$\Delta^t log(p_{it}) \approx \Delta^t log(c_{it}) + \Delta^t \frac{m_{it}}{c_{it}}$

I can then run the following regression across product modules, with store-year fixed effects to absorb rent and labor costs:

$\Delta^t log(p_{it}) = \beta I_i + \lambda_{st} + \epsilon_{it}$

$\Delta^t log(c_{it}) = \widetilde{\beta} I_i + \widetilde{\lambda_{st}} + \widetilde{\epsilon_{it}}$

$\Delta^t \frac{m_{it}}{c_{it}} = \bar{\beta} I_i + \bar{\lambda_{st}} + \bar{\epsilon_{it}}$

with I_i income rank of module. Note that $\beta \approx \widetilde{\beta} + \bar{\beta}$. We can use this relationship[31] to answer the following question: do prices rise slower for high-income consumers because retailer margins decline faster or because wholesale costs rise slower?

As shown in Figure 1.8 and Table 1.8, changes in retailer margins account for 57% of the differential inflation between high- and low-income households. This number can be thought of as a lower bound on the total share of changes in markups in the overall inflation difference, because wholesalers' themselves have a markup. I have checked that this relationship is robust across years and with using other specifications. These results provide strong support for the second prediction of the model: variable markups are a key channel.[32]

Table 1.8.: Changes in Wholesale Costs vs. Changes in Retailer Margins

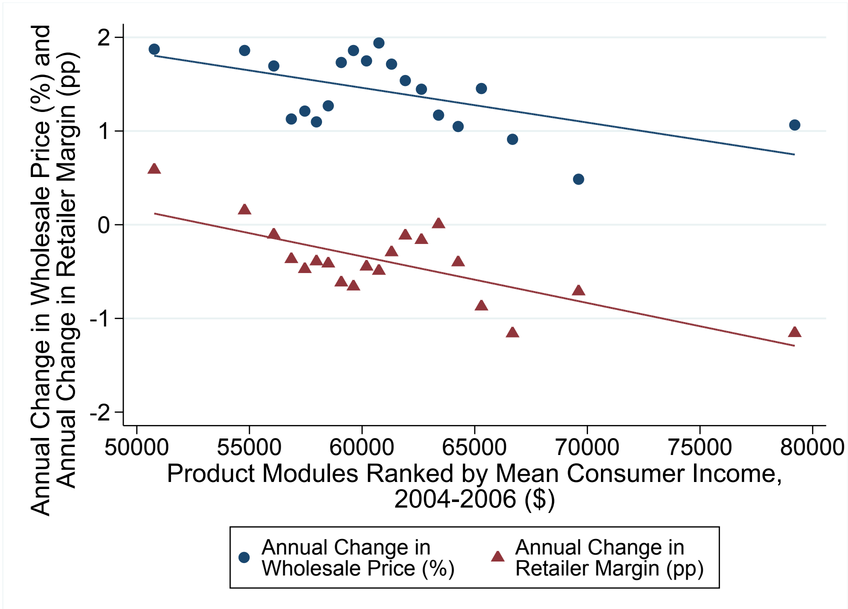
	Log Price Change	Log Wholesale Cost Change	Retailer Margin Change (pp)
$ProductModuleIncomeRank_M$	-0.777***	-0.341***	-0.448***
	(0.188)	(0.103)	(0.212)
Spending Weights	Yes	Yes	Yes
Store-Year Fixed Effects	Yes	Yes	Yes
Number of Observations	6,002,235	6,002,235	6,002,235
Number of Clusters	628	628	628

Standard errors clustered by product modules.

[31] As can be checked from the regression table, the margins are sufficiently small for the Taylor expansion to be almost exact, which in turn implies the teh relationship between the regression coefficients is almost exact.

[32] Variable markups are often studied in the macro literature in the context of short-run business cycle fluctuations. The fact that markups explain a large share of the difference in inflation between high- and low-income households does not mean that these dynamics are bound to be short lived. Indeed, the set of available products changes over time. Adjusted for quality, the marginal cost of the new products is lower than that of existing products, which are forced to reduce their markups. In other words, the price effects show up largely through changes in markups, but these changes reflect the productivity gains brought about by new products.

Figure 1.8.: Changes in Wholesale Costs vs. Changes in Retailer Margins



1.4.4. The Causal Effect of Growing Demand on Product Innovations and Inflation

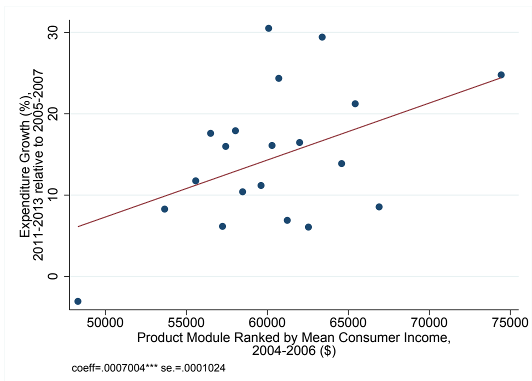
1.4.4.1. Motivating Evidence and Identification Challenge

They key causal channel in the model is that growing demand causes more product entry, and in turn lower prices on existing products due to a fall in markups. Figure 1.9 shows that product modules catering to higher-income households indeed have both higher growth and lower inflation. Moreover, Table 1.4 provided early evidence that supply factors play an important role in differential product introductions across the product space.

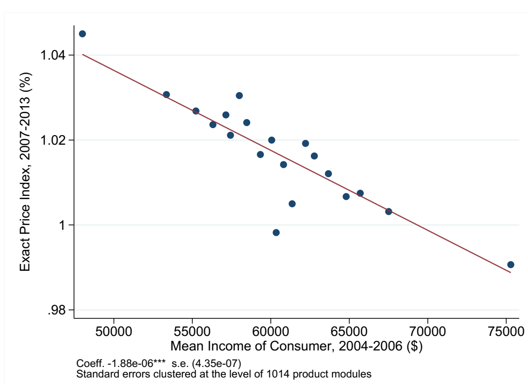
However, these facts alone do not establish that the endogenous response of supply to demand is a key channel. The equilibrium relationship between price and quantity across product modules modules does not identify the causal effect of demand, because of reverse causality (demand might be following supply) and omitted variable bias (there might be unobserved heterogeneity in the difficulty of innovating across modules, which could happen to coincide with spending patterns from non-homothetic preferences). In the remainder of this section, I build a predictor of (potential) demand that is plausibly orthogonal to supply factors. Specifically, I consider changes in market size across product modules over time at the national level driven by changes in the age and income distributions. In robusness checks, I use variation in market size both over time and across local markets within the US.

Figure 1.9.: Product Modules Catering to Higher Income Households Have Faster Growth and Lower Inflation

Panel A: Growth Across Product Modules



Panel B: Inflation Across product Modules



1.4.4.2. Research Design

A major difficulty in any investigation of the impact of market size on innovation is the endogeneity of market size: better products will have larger markets. A strategy to overcome this problem is to exploit variations in market size driven by US demographic changes, which should be exogenous to other, for example scientific, determinants of innovation and entry of new products. To estimate potential market size, I construct age-income profiles of users for each product module $\times$ price decile, and then compute the implied market size from aggregate demographic changes given these (time invariant) income-age profiles. This identification strategy is similar to Acemoglu Linn (2004). Using this strategy, Acemoglu and Linn (2004) showed that large R&D efforts in the pharmaceutical industry endogenously respond to market size. By focusing on product innovations in retail, I study innovation dynamics of a very different nature. More importantly, this paper is the first to examine the causal effect of changes in market size on the price of existing products, as well as on the aggregate price taking into account the welfare gains from increased product variety. I find that prices go down when demand goes up, i.e. the observed supply curve is downward sloping.[33]

The predictor of market size is built as follows. At the beginning of the sample (2004-2006), I compute per capita expenditures $E_{MG}^{T_0}$ in product module $\times$ price decile M for fifteen age-income groups G I consider.[34] Then, I predict (potential) demand at time t as:

$D_{Mt} = \sum_G E_{MG}^{T_0} P_{Gt}$

Thus, the spending profiles are kept consant and the variation in predicted demand comes entirely from changes in age-income group size P_{Gt}. To implement this design, I compute

[33] In ongoing work, I use markups as another outcome to test additional predictions of the model presented at the beginning of this section.

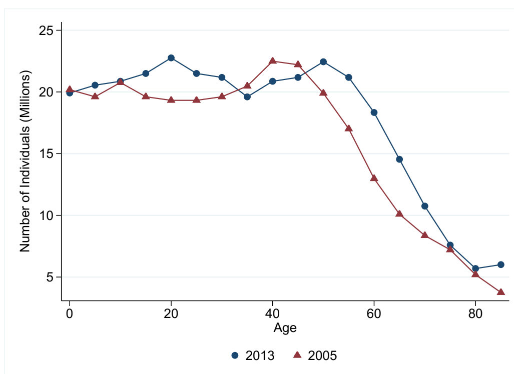
[34] Specifically, I consider the interaction of three age groups - below 45, between 45 and 65, and above 65 - and five income groups - annual household income below 25k, 25k to 45k, 45k to 60k, 60k to 100k, and above 100k.

growth of demand based on the change in the size of each age-income group in 2011-2013 relative to 2004-2006.

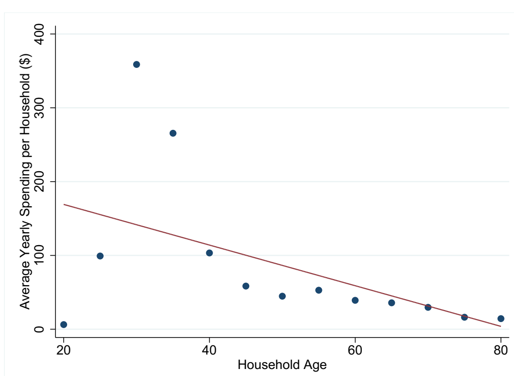
The identification assumption is that the direct effect of changes in age-income group size on the equilibrium price was only through demand. For instance, if the 20-year old were better at creating new products targeting people in the same age group, the identification assumption would be violated. As a robustness check, I repeat the analysis for older households, who are closer to retirement age. I find similar point estimates, which suggests that direct supply effects are not driving the results.

Figure 1.10.: Changes in Market Size from Changes in the Age Distribution

Panel A: Changes in Age Distribution, 2005 to 2013



Panel B: Example of a Product with a Steep Age Profile, Diapers



1.4.4.3. Results

I first present the results with a series of binned scatter plots, where each dot represent 10% of the data. I then show the results in a regression table, with standard errors clustered by product module. Figures 1.11 and 1.12 below show that the predicted increase in market size (based on the changes in the age and income distributions) is positively correlated with the introduction of new products and negatively correlated with inflation. These results lend strong support to the hypothesis that supply endogenously responds to changes in market size.

Figure 1.11.: Higher Market Size Leads to Product Entry

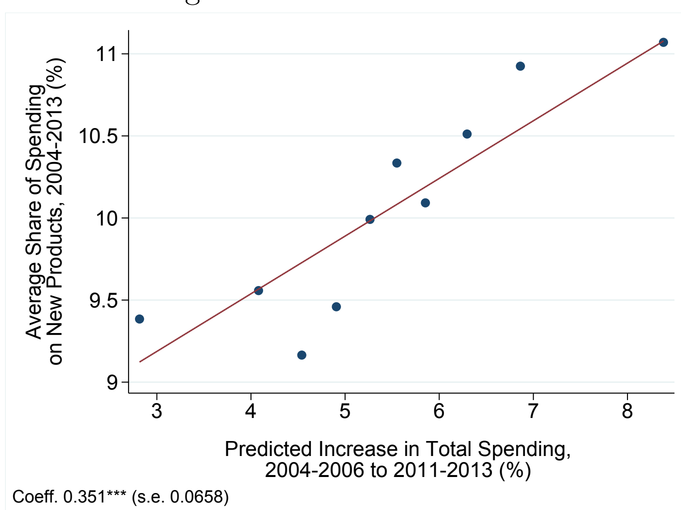


Figure 1.12.: Higher Market Size Leads to Lower Inflation (Overlapping Goods)

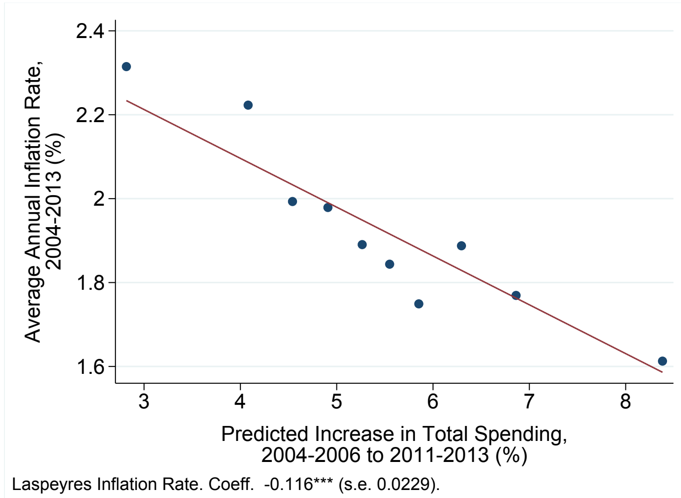


Table 1.9 shows that the relationships between predicted market size growth, product innovations and inflation are significant at the 1% level. The interpretation of the magnitudes is as follows: a one percentage point increase in the growth of demand[35] causes a 0.35 percentage point increase in the share of spending on new products and a 0.11 percentage point decline in the inflation rate on goods that are available across years. Figure A.13 in Appendix D shows the relationship between predicted and actual spending growth, which is also strong. Column 3 of 1.9 confirms that the relationship between predicted and actual growth of total spending is significant at the 5% level.[36]

[35] where growth of demand is measures as the predicted growth in total spending in a product module - price decile given changes in the age and income distributions.

[36] The point estimate is close to 1, i.e. the predictor is unbiased. Unbiased prediction wasn't necessarily expected, because the measure of actualy total spending growth takes into account both price and quantity effects, while the predicted increase in spending is based on the assumption that spending per capita is fixed.

Table 1.9.: Causal Effects of Changes in Market Size

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)	Actual Spending Growth (%)
Predicted Increase in Spending (%)	0.351***	-0.116***	1.031**
	(0.0658)	(0.0229)	(0.492)
Product Module Fixed Effects	Yes	Yes	Yes
Spending Weights	Yes	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes	Yes
Number of Observations	9,089	9,089	9,089
Number of Clusters	1,006	1,006	1,006

Standard errors clustered by product modules.

The point estimates are precisely estimated and can be used for a calibration. From the public use micro data from the US Census, I find that between 2004 and 2013 on average the number of high-income households grew 3.12pp faster than the number of low-income households. Multiplying this number by the point estimate in the second column of Table 1.9 implies an annual inflation difference of 34.3 basis points, which represents 84% of module-decile benchmark[37] and 52% of the overall inflation difference between high- and low-income households. In other words, the response of inflation to changes in market size is sufficiently large to explain most of the difference in inflation rates across the income distribution.

Table 1.10 shows the robustness of these results. Panel A runs a falsification test in the set of product modules - price deciles that experienced negative spending growth during the period 2004-2013. The model does not predict a significant relationship between change in market size and entry of new products or inflation in this subsample, and indeed I do not find any. Panel B addresses the potential concern that some of the relationship between predicted demand and innovation and inflation could be spuriously driven by a differential

[37] The regressions are all at the product module $\times$ price decile level, therefore this is the relevant benchmark.

increase in supply across the product space. It shows that the results are very similar when considering product module - price deciles that cater to consumers above the age of fifty. In other words, the result is not driven by young consumers, for whom direct supply effects are more likely to exist.

Table 1.10.: Robustness of Causal Effects of Changes in Market Size

Panel A: Falsification Test in Product Module - Deciles with Negative Spending Growth

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	-1.093	0.162
	(1.148)	(0.108)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Negative Spending Growth	Yes	Yes
Number of Observations	632	632
Number of Clusters	305	305

Standard errors clustered by product modules.

Panel B: The Effect Is Not Driven by Young Consumers

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	0.306***	-0.113***
	(0.075)	(0.021)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes
Number of Observations	6,571	6,571
Number of Clusters	926	926

Sample restricted to product modules - price deciles with mean consumer age above 50.
Standard errors clustered by product modules.

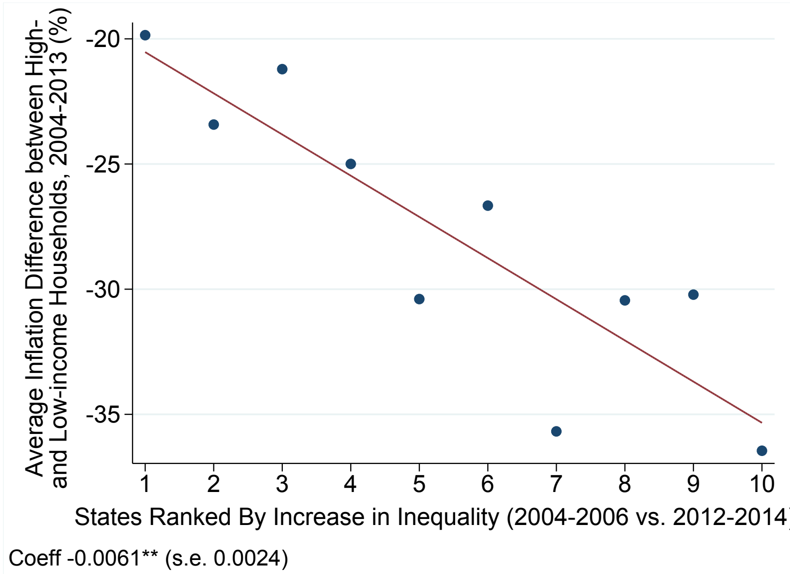
A variety of additional robustness checks are reported in Appendix D. Panel A shows that the points estimates are very stable when including flexible controls for the initial (2004-2006) age and income distributions in each product module - price decile. Specifically, I control linearly for the 10th, 25th, 50th, 75th and 90th percentiles, as well as the mean, of both the age and income distributions. Panel B also shows stability of the point estimates when introducing fixed effect for each price decile within a product module and when omitting product module fixed effects. Panel B of Table A.11 shows that the results are similar when using truncated weights. Statistical significance at the 1% level is maintained in all specifications, with standard errors clustered by product modules.

1.4.4.4. Additional Evidence from Variation Across US States and Local Markets

I provide additional evidence for the relevance of the "market size" hypothesis by exploiting variation in the rate of inequality growth across US states. Using Census public use micro-data between 2004-2006 and 2012-2014, I measure the change in the total income accruing to households who earned more than 100k and less than 30k in each state. Inequality has increased in all 50 states but the rate of increase varied across sates. The increase in inequality was fastest increase in California, Texas and New York and slowest in West Virginia, New Mexico and North Dakota.

UPCs can be thought of as are partly non-tradable because of strength of local brand preferences (Bronnenberg, Dube and Gentzkow, 2012). Therefore, the Nielsen data can be aggregated at the state level to examine how variation in the rate of inequality growth relates to patterns of inflation. In all states, inflation was lower for high-income households earning above $100,000 a year, relative to low-income households making below $30,000 a year. But this difference in inflation rates was relatively larger in states with a faster increase in inequality. Figure 1.13 shows this result.

Figure 1.13.: The Inflation Difference Between High- and Low-Income Increases as Inequality Increases Faster



In Appendix D, as a robustness check I use time variation in the household age and income distributions in seventy-six local markets tracked by Nielsen within the US between 2004 and 2013. I compare inflation patterns across product module - local market cells with increasing or decreasing predicted market size. I again find that inflation is lower when predicted demand increases - the point estimates are very similar to those obtained from the analysis at the national level and robust to the inclusion of various fixed effects.

1.4.5. Alternative Mechanisms

I have investigated various alternative explanations for the evidence. First, I show that broad shocks cannot explain the results (shock at the level of MSAs, product groups; recession; rise of online retail; oil shocks). Second, I examine the hypothesis that the production function features increasing returns to scale, using retailer marginal cost data for a subset of the products in 2004-2007. Third, I study whether dynamic pricing is consistent with the evidence. Finally, I investigate whether trade patterns can explain the results.

Aggregate shocks. First, the various decompositions reported in Section 1.3 show that the results are not driven by broad shocks that would be specific to certain areas (Appendix Table A.10) or to certain departments, product groups or product modules (Tables 1.1 and 1.2).

Online retail. The rise of online retail could have differentially benefited high- and low-income households. For instance, if higher-income households are more technology savvy, they might be more likely to use online platforms to search for products, which would increase their price elasticity and result in lower equilibrium markups. However, the inflation difference across product categories is not related to heterogeneity in exposure to online retail - in particular, it persists in categories that were very little affected by online retail during this period, such as food (Table 1.1).

Innovation dynamics independent of changes in market size. An alternative view of the innovation patterns is that product innovation may always be skewed towards the higher-income consumers, regardless of the underlying patterns of growing inequality. In other words, the patterns documented in Section 1.3 may be a steady state By introducing flexible controls for the income distribution of consumers and for the quality distribution (price deciles) within a product module, Panel B of Appendix Table A.11 shows that the estimated response of product innovations ot market size is not confounded by static patterns related to income or quality. Moreover, I have not found empirical support for the predictions of a simple class of models that generate a steady-state difference in the inflation rates experienced by high- and low-income households - in these models, the equilibrium price elasticity of higher-income consumers should always be lower.[38]

[38] Intuitively, if high-income consumers are less price elastic and if the cost of increasing product variety is linear, in equilibrium we will observe a high flow of new products targeting higher income consumers.

Manufacturer competition. Using the manufacturer identifiers provided by GS1, I have studied patterns of market concentration at the manufacturer level and how they have changed over time. Concentration (as measured by Herfindahl indices) tends to be lower in parts of the product space catering to higher-income households, but changes in concentration over time are small. Overall, the data shows that entry of new manufacturers is not the driving force of the results. This justifies focusing on a model where changes in markups are induced by changes in the elasticity of market demand, rather than by competition between oligopolists, e.g. in a Cournot model.

Household search behavior. Another possible channel for the results is that high-income consumers could have become more price elastic because their search behavior has changed. Such a channel would manifest itself primarily through within-UPC inflation difference between high- and low-income households, which Table 1.2 shows is not the case.

Other mechanisms. In ongoing work, I use Nielsen TDLink data to characterize changes in the competitive environment of retailers and I document how competitive dynamics differ across areas depending on the density of high- and low- income households. I also examine whether trade is an important channel by studying heterogeneity in the inflation difference across product groups with different degrees of import penetration. Finally, I test the predictions of models featuring dynamic pricing and increasing returns to scale.

1.5. Conclusion

In this paper, I have shown that quality-adjusted inflation substantially varies across the income distribution in the retail sector. The current methodology of statistical agencies like BLS cannot capture this variation, which exists primarily at the product level rather than

The equilibrium mechanism is that the high-end products have higher margins (because the high-income consumers are less price elastic) but have a shorter lifecycle (because they get displaced by other high-end product innovations).

across broad item categories. Furthermore, I have established that product introductions and prices endogenously respond to changes in market size in a way that magnifies the welfare effects of changes in nominal inequality. As shown in a simple calibration, the endogenous response of supply to changes in market size over the past decade can explain most of the observed difference in inflation rates across the income distribution during this period. These findings open up several directions for future research. Do similar results hold beyond the retail sector? How should one adjust optimal redistributive taxation formulas (e.g. as in Mirrlees, 1971) to take into account the endogenous response of supply to changes in market size? These and other extensions await further research.

2. Team-Specific Capital and Innovation[1]

2.1. Introduction

Teamwork has become an essential feature of modern economies and knowledge production (Seaborn, 1979, Wuchty et al., 2007, Jones, 2010, Crescenzi et al., 2015, and Jaffe and Jones, 2015). We investigate empirically the importance of *team-specific capital* for the compensation and patent production of inventors, using administrative tax and patent data for the population of US patent inventors from 1996 to 2012. While general human capital augments productivity at all firms (Becker, 1975), and while firm-specific capital augments productivity with any existing or future collaborators within the firm (Topel, 1991), team-specific capital makes an inventor more productive with their existing co-inventors. Team-specific capital encompasses skills, experiences and knowledge that are useful only in the context of a specific collaborative relationship: high team-specific capital means that the collaborative dynamics in the team are unique and difficult to rebuild with other collaborators, which improves each inventor's ability to produce valuable innovations with these specific co-inventors.[2] If the collaboration between two patent inventors were to exogenously end,

[1] Co-authored with Neviana Petkova and Alex Bell.

[2] Team-specific capital can result from a "match" component which is constant over time, for instance if two inventors are a particularly good fit for each other, or from an "experience" component which increases the

would this have a significant and long-lasting impact on the career, compensation, and productivity of these inventors? Or are co-inventors easily substituted for, beyond short-term disruption of ongoing work? In other words, is team-specific capital an important ingredient of the typical inventor's lifecycle earnings and productivity, much like firm-specific capital is crucial for the typical worker? This paper establishes the existence and economic relevance of patent inventors' team-specific capital.

We provide causal estimates of what the typical inventor would lose, in terms of labor earnings, total earnings and patent production, if a collaboration with one of their co-inventors were to end exogenously. Using a detailed merged dataset of United States Patent and Trademark Office (USPTO) patents data and Treasury administrative tax data, we use the premature deaths of 4,714 inventors, defined as deaths that occur before the age of 60, as a source of exogenous variation in collaborative networks. The causal effect is identified in a difference-in-differences research design, using a control group of patent inventors whose co-inventors did not pass away but who are otherwise similar to the inventors who experienced the premature death of a co-inventor. We find that ending a collaboration causes a large and long-lasting decline in an inventor's labor earnings (- 3.8% after 8 years), total earnings (- 4% after 8 years) and citation-weighted patents (- 15% after 8 years). This evidence implies that the continuation of collaborative relationships has substantial specific value for the typical inventor, approximately equal to half of the returns to one year of schooling (Mincer, 1973). It rejects the alternative hypothesis that continued collaborations are not a key ingredient in an inventor's earnings function and patent production function beyond short-term disruption of ongoing work.

To establish team-specific capital as the primary explanatory mechanism, we show that the

value of the collaboration over time, for example if two inventors learn how to best collaborate with each other over the course of several joint projects. Appendix C discusses the extent to which our results can help distinguish between the "match" and "experience" components of team-specific capital. The evidence can only be suggestive, because we do not have random variation for the timing of formation of the collaborations.

gradual decline in earnings and citation-weighted patents following the premature death of a co-inventor is driven by the fact that the inventor lost a partner with whom they were collaborating extensively, which made additional co-inventions impossible. We do so in four steps. First, we rule out alternative explanatory mechanisms that are not specific to the team. We establish that the effect does not stem from the disruption of the firm or from network effects by estimating the causal effect of an inventor's death on their coworkers and on inventors that are two nodes away from the deceased in the co-inventor network.[3] Second, we show that the effect is not driven by top-down spillovers from unusually high-achieving deceased inventors (e.g. as in Azoulay et al., 2010, and Oettl, 2012). Third, we demonstrate that the intensity of the collaboration between an inventor and their deceased co-inventor prior to death is an important predictor of the magnitude of the effect. Fourth, we document that the effect of co-inventor death on an inventor's patents is much smaller when patents that were co-invented with the deceased are not taken into account in the difference-in-differences analysis: although the survivor's own patents suffer as well, the effect primarily applies to co-invention activities with the deceased. We also show that team-specific capital matters in all technology categories, at various levels of the distribution of patent quality, and spans firm and geographic boundaries. In Section IV, we discuss whether other mechanisms could be consistent with the evidence.

Beyond establishing the first-order importance of team-specific capital, the paper makes two additional contributions. First, we present new descriptive statistics on collaboration patterns and the composition of teams. We find that assortative matching is true only up to a point: there is wide variation in the relative earnings and age of co-inventors. Second, we introduce a novel specification to estimate the causal effect of an individual's premature death, which includes all leads and lags around co-inventor death in both the treatment and

[3] In addition to ruling out important alternative mechanisms that could explain our finding, this analysis yields new insights about substitution and complementarity patterns between inventors in the innovation production function. See Section IV for a complete discussion.

control groups. This specification is robust to mechanical statistical patterns induced by the construction of the sample, which have not been addressed in the existing literature and which we show result in substantial biases of the estimates of interest.[4]

Our work relates to several strands of literature. The use of premature deaths as a source of identification is becoming increasingly common (Jones and Olken, 2005, Bennedsen et al., 2007, Azoulay et al., 2010, Nguyen and Nielsen, 2010, Oettl, 2012, Becker and Hvide, 2013, Fadlon and Nielsen, 2014, Isen, 2015) and several papers have investigated peer effects in specific areas of science (Azoulay et al., 2010,[5] Borjas and Doran, 2012, 2014, Oettl, 2012, Waldinger, 2010, 2012). In contrast, our paper studies peer effects among patent inventors in all technology classes, using both earnings and patent data (Moser et al., 2014, study German emigres' effects on US chemical patents). We estimate the differential spillover effect of an inventor on various peer groups (co-inventors, coworkers, and second-degree connections in the co-inventor networks) using the same research design, which allows us to establish the unique importance of co-inventors in an inventor's career. Other related strands of literature study the role of teams in innovation (e.g. De Dreu, 2005, Jones, 2009, Agrawal et al., 2013,

[4]It is not sufficient to control for age, year and individual fixed effects in the difference-in-differences estimator, because these fixed effects do not fully account for the trends in lifecycle earnings and patents around the year of co-inventor death. Intuitively, an inventor must necessarily have invented a patent before the year of death of their co-inventor and is more likely to have been employed at that time, even conditional on a large set of fixed effects. We show that this results in a substantial bias in the estimate of the causal effect for several of the outcomes we study in this paper. Including a full set of leads and lags around co-inventor death for both treated and control inventors addresses this problem. This solution is an application of the standard difference-in-differences estimator, where treatment occurs at only one point in time, to our setting, where co-inventor deaths are scattered across years. Similar considerations apply when estimating heterogeneity in the treatment effect. See Section III and Appendix D for more details and for a comparison with the existing literature using premature death research designs.

[5]Our paper is closest to Azoulay et al. (2010), who examine the effect of the premature deaths of 112 eminent life scientists on their coauthors. We build on their groundbreaking identification strategy and advance the literature in several respects. First, we characterize peer effects using earnings data in the full population of patent inventors and we not only examine the effect on co-inventors but also on co-workers and inventors who are two nodes away from each other in the co-inventor network. Second, our findings are fundamentally different: whereas Azoulay et al. (2010) show that "star scientists" are irreplaceable in a way that is not related to team dynamics, we find an effect of the premature death of co-inventors that are not "superstars," and this effect is much larger for more closely-knit teams and is driven by co-invention activities (see Section IV for a complete discussion).

Alexander and van Knippenberg, 2014), examine the notion of team-specific or network-specific human capital from a theoretial perspective (e.g. Mailath and Postlewaite, 1990, Chillemi and Gui, 1997), investigate the effect of co-mobility of colleagues (Hayes et al., 2006, Groysberg and Lee, 2009, Campbell et al., 2014) and develop theories of knowledge spillovers across inventors (e.g. Stein, 2008, Lucas and Moll, 2014). As discussed in Section IV, our results on spillover effects between inventors who are two nodes away from each other in the co-inventor network provide a unique test of competing models of strategic interactions in networks (Jackson and Wolinsky, 1996, Bramoulle et al., 2014). Finally, this paper is part of a nascent literature using administrative data to describe the careers of patent inventors (Toivanen and Vaananen, 2012, Bell et al., 2015, Dorner et al., 2015, Depalo and Di Addario, 2015).

This paper has important implications for innovation policy. Our findings suggest that investing in improving the match technology between inventors and encouraging the accumulation of team-specific capital could lead to substantial productivity gains. Furthermore, our results indicate that research teams are an important vehicle for knowledge transmission and such team collaborations affect the productivity of team members outside of their joint projects (even though, as noted earlier, the effect is larger for co-invention activities). Teamwork improves inventors' productivity and increases their incomes, which generates additional tax revenues and creates large fiscal externalities.

The remainder of the paper is organized as follows. In Section II, we present the dataset and novel descriptive statistics on the composition of teams. In Section III, we describe the research design and present the estimates of the causal effect of the premature death of a co-inventor on an inventor's compensation and patents. In Section IV, we distinguish between various mechanisms. Section V concludes. Several robustness checks, heterogeneity results and empirical estimation details are deferred to the Appendices. Appendix A reports additional summary statistics and tests for balance between treated and control groups.

Appendix B presents robustness checks on the causal effect of co-inventor death. Appendix C conducts additional tests for heterogeneity in the effect of co-inventor death. Appendix D provides additional details on our econometric framework. Appendix E describes the construction of the dataset and reports additional summary statistics on the composition of inventor teams.

2.2. Data and Descriptive Statistics

2.2.1. Data Construction

We use a merged dataset of United States Patent and Trademark Office (USPTO) patents data and Treasury administrative tax files as in Bell et al. (2015). The patent data are extracted from the weekly text and XML files of patent grant recordations hosted by Google. The raw files contain the full text of about 5 million patents granted from 1976 to today, extracted from the USPTO internal databases in weekly increments.

Administrative data on the universe of U.S. taxpayers is sourced from Treasury administrative tax files. We extract information on inventors' city and state of residence, wages, employer ID, adjusted gross income, as well as current citizenship status and gender from Social Security records. Most data are available starting in 1996, however wages and employer ID are available only starting in 1999, which marks the beginning of W-2 reporting. Inventors from the USPTO patent data are matched to individual taxpayers using information on name, city and state of residence (Appendix A describes the iterative stages of the match algorithm). The match rate is over 85% and the matched and unmatched inventors appear similar on observables, as documented in Bell et al. (2015). Any inventor with a non-U.S. address in the USPTO patent data is excluded from the matching process and dropped from the sample. The resulting dataset is a panel of the universe of U.S.-based inventors, tracking over 750,000 inventors from 1996 to 2012. The employer ID is based on the Employer Identification Number (EIN)) reported on W-2 forms. We show in Appendix Figure E1 that the distribution of EIN size is very similar to the distribution of firm size in the Census. In the rest of the paper, we refer to business entities with distinct EINs as distinct firms.[6]

2.2.2. Identifying Deceased Inventors, Survivor Co-inventors, Second-Degree Connections and Coworkers

We construct various groups of inventors to carry out the premature death research design. We start by identifying 4,924 inventors who passed away before the age of 60 and were granted a patent by USPTO before their death.[7] Information on the year of death and age at death is known from Social Security records. The cause of death is not known. In order to reduce the likelihood that death results from a lingering health condition, we consider inventors passing away before 60 and, in robustness checks, we repeat the analysis by excluding deceased inventors who ever claimed tax deductions for high medical expenses.

We construct a group of "placebo deceased inventors" who appear similar to the prematurely deceased inventors but did not pass away. Specifically, we use a one-to-one exact matching procedure on year of birth, cumulative number of patent applications at the time of (real or placebo) death, and year of (real or placebo) death in order to identify placebo deceased inventors among the full population of inventors.[8] 4,714 deceased inventors find an exact

[6] In some cases, it could be that business entities with different EINs are the subsidiary of the same parent company. However, treating distinct EINs as distinct firms is standard practice in the literature (Song et al., 2015).

[7] As described below, ultimately we analyze only 4,714 premature deaths due to the lack of appropriate matches for the remaining prematurely deceased inventors. We consider prematurely deceased inventor who are weakly below 60, i.e. we keep inventors who are 60 in the year of death.

[8] The match is conducted year by year. For instance, for inventors who passed away in 2000, we look for exact matches in the full sample of inventors - an exact match is found if the control inventor was born in the same year and had the same number of cumulative patent applications as the deceased in 2000. The inventors

match using this procedure.[9] Thus, we obtain a control group of placebo deceased inventors who have exactly the same age, the same number of cumulative patent applications and exactly the same year of (placebo) death as their associated (real) deceased inventor.

Next, we build the co-inventor networks of the real and placebo deceased inventors. Any inventor who ever appeared on a patent with a real or placebo deceased inventor before the time of (real or placebo) death is included in these networks. In the rest of the paper, we refer to these inventors as real and placebo "survivor inventors." We exclude survivor inventors who are linked to more than one real or placebo deceased inventor.[10] We thus obtain 14,150 real survivor inventors and 13,350 placebo survivor inventors. These inventors constitute the main sample used for the analysis carried out in the rest of the paper. Note that we perform the matching procedure on the real and placebo deceased inventors rather than on the survivor inventors - the benefits of this approach are discussed in Section III.

We construct two other groups of inventors, which will be used to differentiate between mechanisms. First, we build the network of inventors who are two nodes away from the real and placebo deceased inventors in the co-inventor network. These inventors are direct co-inventors of the deceased's direct co-inventors, but they never co-invented a patent with any of the (real or placebo) deceased inventors. To increase the likelihood that these inventors were never directly in contact with the deceased, we impose two additional restrictions: of the inventors who are two nodes away from the deceased in the co-inventor network, we keep only those who never worked for the same employer and never lived in the same commuting zone as the deceased inventor. We refer to these inventors as real and placebo "second-

from the full sample that match are then taken out of the sample of potential matches, and the procedure is repeated for the following year, until the end of the sample. This matching procedure without replacement thus determines a counterfactual timing of death for the placebo deceased inventors. When there is more than one exact match, the ties are broken at random.

[9] The 5% unmatched deceased inventors do not significantly differ on observable characteristics from those who find a match, except that they tend to have more cumulative applications at the time of death. In robustness checks presented in Appendix E, we repeat the analysis with a propensity-score reweighting approach which uses data on all deceased inventors and obtain similar results.

[10] We lose only 36 survivor inventors by imposing this restriction.

degree connections" in the remainder of the paper. As before, we exclude inventors in this group who are linked to more than one real or placebo deceased inventors. This procedure yields 11,264 real second-degree connections and 12,047 placebo second-degree connections. Second, we construct the group of "coworkers" of the deceased by identifying all inventors who worked for the same employer as the deceased in the year before death, as indicated on W-2 forms. We exclude coworkers that ever co-invented with a prematurely deceased inventor or who experienced multiple premature death events. Focusing on coworkers in firms with less then 2,000 employees, the final sample consists of 13,828 real coworkers and 14,364 placebo coworkers.[11]

2.2.3. Variable Definitions and Summary Statistics

In the analysis carried out in the rest of the paper, we study various outcome variables at the individual level from 1999 until 2012. First, we consider inventors' labor earnings, which refer to annual W-2 earnings. When an inventor does not receive a W-2 form after 1999, we impute their labor earnings in that year to be zero. Second, we construct a measure of an inventors' total earnings, defined as an inventors' adjusted gross income (earnings reported on IRS tax form 1040) minus the W-2 earnings of the inventor's spouse. Adjusted gross income is a tax concept offering a comprehensive measure of a household's income, including royalties, self-employment income and any other source of income reported on 1040 tax forms.[12] We define non-labor earnings as the difference between total earnings and labor earnings. All

[11] We focus on smaller firms to increase the chances that we find a negative effect of an inventor's death on their coworkers, since we are interested in testing whether the effect we document for co-inventors is driven by the disruption of the firm. In Appendix E, we carry out the analysis on the full sample of coworkers, composed of 173,128 real survivor coworkers and 143,646 placebo survivor coworkers, and we find similar results. The difference in the size of the groups of real and placebo coworkers in the full sample is driven by a thin tail of deceased inventors working in firms employing thousands of other inventors, as documented in Appendix Table A5.

[12] A limitation of our measure of total earnings for inventors filing jointly is that we can only subtract the inventor's spouse's W-2 earnings from the household's adjusted gross income, not the spouse's other sources of income, which are unobserved. But the exact same procedure is applied to all inventors in the various groups we consider.

earnings variables are winsorized at the 1% level.[13] Third, we use adjusted forward citations, which are defined for year t as the total number of forward citations received on all patents the individual applied for in year t, divided by the number of inventors who appear on each patent. Forward citations include all citations of the patent made as of December 2012 and are a measure of the "quality" of innovative output. We divide forward citations by the total number of inventors on the patent to reflect the fact that a single inventor's contribution is smaller in larger teams.[14] Fourth, we use the number of patents granted by the USPTO as of December 2012, as well as the number of patents in the top 5% of the citation distribution.[15] Lastly, we create indicator variables that turn to one when labor earnings are greater than 0 or above thresholds for the 25th, 50th and 75th percentiles of the labor earnings distribution.[16] We proceed similarly for total earnings. These indicators are used as outcome variables to characterize the effect of an inventor's premature death on their co-inventors' compensation at different quantiles of the income distribution. Since labor earnings are only available from 1999 onwards, for consistency we do not use data prior to 1999 for any of the variables in the analysis, but the results are qualitatively similar when pre-1999 data is included for adjsuted gross income, patent applications and citations.

Table 1 presents summary statistics for the variables of interest in the main samples used in the analysis. Statistics on total earnings and wages are computed based on the entire panel for the full sample of inventors, and based on years before the death event for the

[13]We have checked that the results are robust to winsorizing at the 5% level

[14]This is common practice. We check the robustness of our results with other measures of citations, which do not adjust for team size, take into account citations only over a fixed rolling window of a couple years around application or grant (in order to address censoring issues), and distinguish between examiner-added and applicant-added citations. Section III discusses these various robustness checks.

[15]We define the count of patents in the top 5% of citations as the number of patents the survivor inventor applied for in a given year that were in the top 5% of the citation distribution, where the distribution is computed based on all patents that were cited, applied for in the same year and in the same technology class (we aggregate USPC classes into six main technology classes, as is common in the literature). Throughout the paper, we consider only patents that were granted as of December 2012 and we use the year of filing of the patent application as the year of production of the invention.

[16]These quantiles are computed before the time of death in the population of real and placebo survivor inventors.

deceased and the survivor inventors. Age, cumulative applications and cumulative citations are computed in the year of death for the deceased and the survivors, and across all years for the full sample. Appendix Table A3 presents similar statistics for the second-degree connections and coworkers.

The real deceased inventors are on average seven years older than inventors in the full sample. By construction, the distribution of age at death for the placebo deceased inventors exactly matches that of the real deceased inventors. Likewise, the distribution of the number of applications is the same for real and placebo deceased inventors. The distribution of labor earnings, total earnings and forward citations is also very similar in these two groups, although our matching algorithm did not match on these variables.

The real and placebo survivor inventors are also older than inventors in the full sample and they have much higher labor earnings and total earnings and many more patent applications and citations. The age difference is due to the fact that there is assortative matching by age in inventor teams, as documented in Section II.D, and the deceased are older than inventors in the full sample. The difference in compensation and patents is due to a selection effect: inventors who have co-invented many patents are more likely to experience the (real or placebo) death of one of their co-inventors. Therefore, it would not be appropriate to use the full population of inventors as a control group for the real survivor inventors, as their lifecycle earnings are likely to be on different trajectories. In contrast, the distributions of labor earnings, total earnings, age and patent applications and citations are very similar in the group of placebo survivors and real survivors. Importantly, our matching algorithm did not impose that any of the characteristics of the placebo survivor inventors should be aligned with those of the real survivor inventors, since we matched on characteristics of the real and placebo deceased only. Labor earnings are slightly lower for the real survivors compared to the placebo survivors, but we will check in Section III that this difference is constant during years prior to co-inventor death, consistent with the assumptions of the difference-in-differences research design. Appendix Tables A1 and A2 show that the real and placebo survivors are also similar in terms of the year of co-inventor death, their technology class specialization, the size of their co-inventor networks and the size of their firms.

2.2.4. Descriptive Statistics on Patent Inventor Teams

Most inventors work in teams: 55% of the 1,375,587 patents in our data are produced by teams, i.e. more than one inventor is listed on the patent. Moreover, team composition shows a significant degree of persistence. In our sample, considering teams that applied for a patent in 2002, the probability that another patent applied for by a member of the team between 1997 and 2007 also includes at least one other member of the 2002 team is 30.4%. When conditioning on patents that were assigned to different assignees[17], the percentage falls but remains high, at 21.6%. This suggests that teams are persistent across firm boundaries.[18]

There is wide variation in the composition of inventor teams. Taking teams of two inventors in 2002 as an example, Figure 1 shows the distribution of absolute differences between team members in total earnings, labor earnings, and age. The mean age difference between inventors in these teams is 10, with a standard deviation of 15. In one-fourth of these teams, the age difference is three years or less and the difference in labor earnings is below $25,000. But in another fourth of these teams, the age difference is larger than 14 years and the difference in labor earnings is above $120,000. Therefore, it is true that inventors who are similar in characteristics like age and compensation tend to work together, but only up to a point. Appendix E reports additional results and the findings are qualitatively similar when considering other years and larger teams.

[17] Assignees are the legal patent holders and are typically the employers of the inventors on the patents.

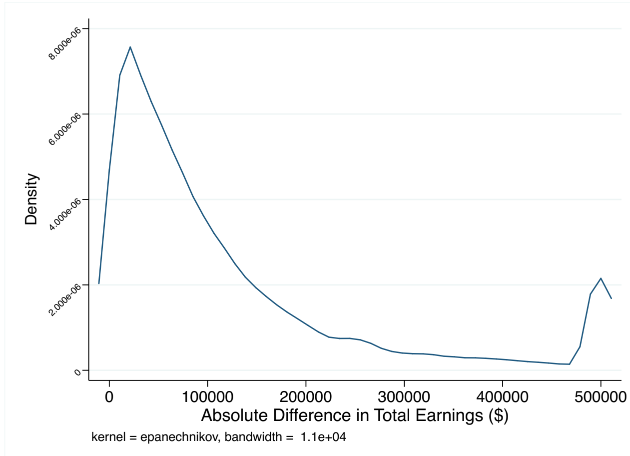
[18] Similar results are obtained when considering other application years as the year of reference. Appendix Table E6 documents that many teams span more than one EIN, which means they most likely cross firm boundaries.

Table 2.1.: Summary Statistics

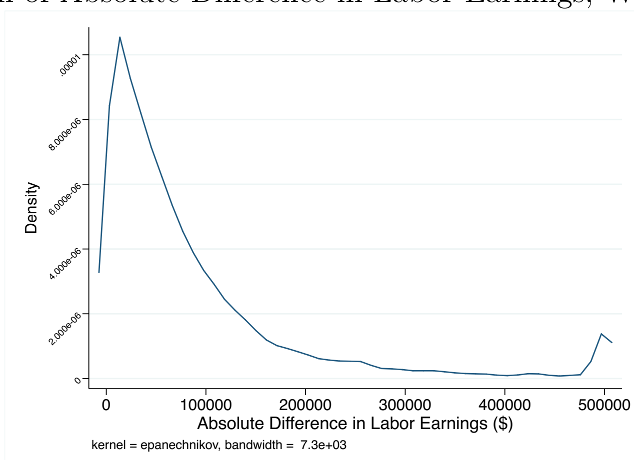
Variable	Sample	Mean	SD	10pc	25pc	50pc	75pc	90pc
	Full Sample	144,096	316,636	38,000	58,000	110,000	163,000	241,000
	Real Deceased	139,857	308,000	35,000	59,000	105,000	160,000	237,000
Total Earnings	Placebo Deceased	139,102	320,970	36,000	58,000	104,000	162,000	236,000
	Real Survivors	177,020	355,347	48,000	89,000	125,000	173,000	270,000
	Placebo Survivors	177,247	360,780	47,000	89,000	125,000	173,000	271,000
	Full Sample	117,559	257,466	25,000	46,000	90,000	142,000	202,000
	Real Deceased	121,691	258,289	29,000	50,000	99,000	147,000	210,000
Labor Earnings	Placebo Deceased	124,149	248,546	33,000	52,000	101,000	148,000	210,000
	Real Survivors	152,602	295,832	42,000	78,000	113,000	160,000	239,000
	Placebo Survivors	155,098	290,201	44,000	80,000	116,000	162,000	242,000
	Full Sample	2.31	2.51	0	1	1	3	7
	Real Deceased	2.50	2.43	0	1	1	3	7
Cumulative Applications	Placebo Deceased	2.50	2.43	0	1	1	3	7
	Real Survivors	12.42	28.31	1	2	5	13	28
	Placebo Survivors	11.92	29.52	1	2	5	13	27
	Full Sample	6.64	12.2	0	0	1	6.58	23.5
	Real Deceased	8.74	13.09	0	0	3	10	29.13
Cumulative Citations	Placebo Deceased	8.51	13.20	0	0	2.5	9.95	30
	Real Survivors	42.00	171.03	0.25	1.3	7	28.5	89.53
	Placebo Survivors	40.20	164.20	0.32	1.5	7	29.5	85.32
	Full Sample	43.29	9.65	30	36	44	51	56
	Real Deceased	50.85	7.44	40	46	52	57	59
Age	Placebo Deceased	50.85	7.44	40	46	52	57	59
	Real Survivors	47.53	10.89	35	41	48	55	61
	Placebo Survivors	47.289	11.16	34	41	47	55	60
	Full Sample	756,118						
	Real Deceased	4,714						
# Inventors	Placebo Deceased	4,714						
	Real Survivors	14,150						
	Placebo Survivors	13,350						

Figure 2.1.: Team Composition for Two-Inventor Teams in 2002

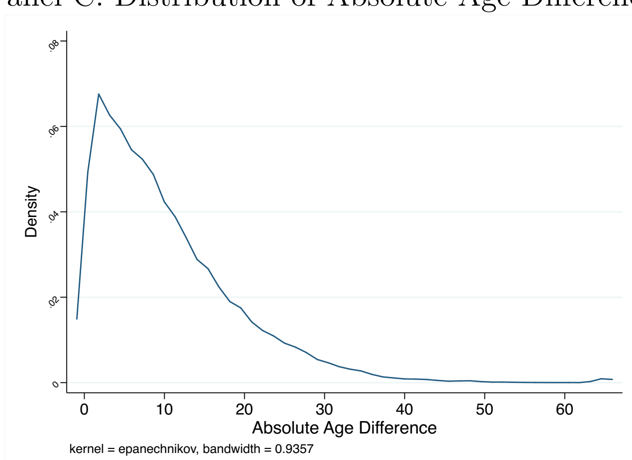
Panel A: Distribution of Absolute Difference in Total Earnings, Winsorized at $500,000



Panel B: Distribution of Absolute Difference in Labor Earnings, Winsorized at $500,000



Panel C: Distribution of Absolute Age Difference



2.3. Estimating the Causal Effect of the Premature Death of a Co-Inventor on an Inventor's Compensation and Patents

This section presents our methodology to estimate the average treatment effect of experiencing death of a coauthor on labor earnings, total earnings, patents and citation-weighted patents. It then describes our main results and a series of robustness checks.

2.3.1. Research Design

We want to build the counterfactual of compensation and patent production for (real) survivor inventors, had they not experienced the premature death of a co-inventor. Two main challenges arise to identify this causal effect. First, the real survivor inventors are on a different earnings and patent trajectory than the full population of inventors. To address this challenge, we use the control group of placebo survivor inventors described in Section II in a difference-in-differences research design. Second, death may not be exogenous to collaboration patterns.[19] We show that the estimated causal effects of co-inventor death are significant only after the year of death, which alleviates this concern.

Figure 2 confirms non-parametrically that the real and placebo survivor inventors are on similar earnings and patent trajectories before the time of co-inventor death and sharply

[19] We cannot think of very convincing examples of why this could be the case, but perhaps a particularly bad collaboration may result in an inventor's death. For a discussion of how pre-trends can be interpreted as anticipation rather than endogeneity of treatment, see Malani and Reif (2015).

differ afterwards.[20] This bolsters the validity of the research design, especially given that our match algorithm did not use any information on survivor inventors. Real and placebo survivors have similar levels of total earnings before death, but placebo survivors have higher labor earnings than the real survivors before death, indicating that real survivors have a higher share of their total earnings in the form of non-labor earnings . The difference in labor earnings appears roughly constant, at around $2,500 (about 2% of labor earnings). In our regression framework, we use individual fixed effects to absorb this difference.

Figure 2 shows that the earnings profile of survivor inventors flattens out after the time of death, even for the placebo survivor inventors. This may be due to curvature in the age profile of earnings, year fixed effects, or mechanical effects induced by the construction of the sample of survivors. Citations are declining over time, probably primarily due to censoring (patents applied for and granted near the end of our sample do not have the opportunity of being cited). Our regression framework takes all of these effects into account.

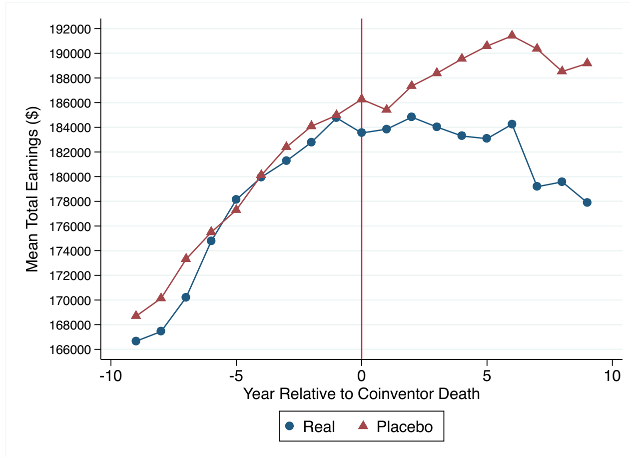
Figure 2 offers a transparent depiction of the data and is useful in gauging the magnitude of the causal effect of co-inventor death on total earnings, labor earnings and forward adjusted citations. However, it is not well suited to a precise estimation of the causal effect - since covariates like age are not perfectly balanced across treated and control groups - nor to robust inference. Two types of clusters are important to take into account for inference: even after controlling for a battery of fixed effects, there may be serial correlation in an inventor's outcomes over time and the outcomes of inventors linked to the same deceased may be correlated. We cluster standard errors at the level of the deceased inventors, which takes into account both forms of clustering.[21]

[20]The figure plots the raw data, without imposing that mean outcomes in the treatment and control groups should be equal prior to death.

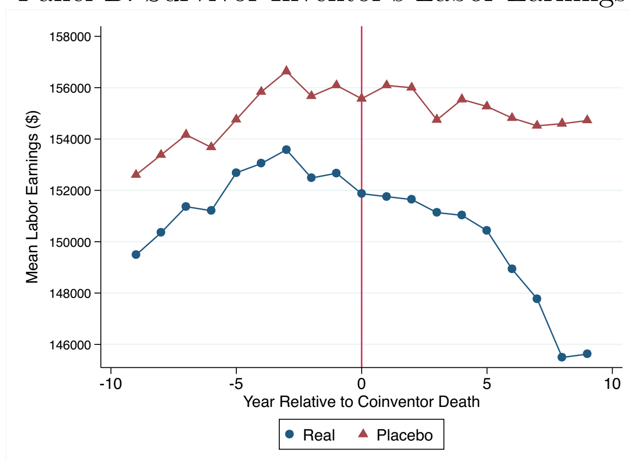
[21]We are close to observing the population of patent inventors who passed away prematurely between 1996 and 2012. Therefore, we interpret our standard errors with respect to their superpopulation. In Appendix Table B10, we use the coupled bootstrap procedure of Abadie and Spiess (2015) to estimate standard errors taking into account the matching step.

Figure 2.2.: Path of Outcomes around Co-inventor Death

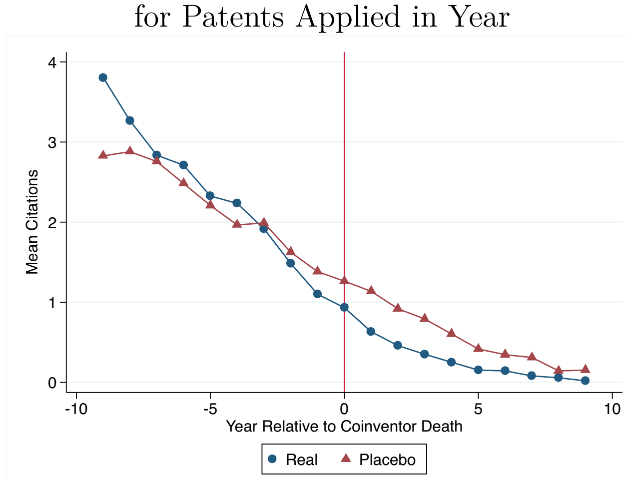
Panel A: Survivor Inventor's Total Earnings



Panel B: Survivor Inventor's Labor Earnings



Panel C: Survivor Inventor's Adjusted Forward Citations Received for Patents Applied in Year



2.3.2. Regression Framework

In order to study the dynamics of the effect, while at the same time probing the validity of the research design by testing whether there appears to be any effect of losing a co-inventor before the event actually occurs, we use a panel data model based on five elements, whose relevance has been discussed in the previous subsection. First, we include a full set of leads and lags around co-inventor death for real survivor inventors (L_{it}^{Real}). The predictive effects associated with these leads and lags are denoted $\{\beta^{Real}(k)\}_{k=-9}^{9}$, where k denotes time relative to death.[22] If the identification assumption described below holds, $\beta^{Real}(k)$ denotes the causal effect of co-inventor death on the outcome of interest k years after death. Second, we use a full set of leads and lags around co-inventor death that is common to both real and placebo survivors (L_{it}^{All}) - the corresponding predictive effects are denoted $\{\beta^{All}(k)\}_{k=-9}^{9}$. Lastly, we introduce three distinct sets of fixed effects: age fixed effects (a_{it}), year fixed effects (γ_t) and individual fixed effects (α_i).

We assume separability[23] and specify the conditional expectation functions as follows:

$E[Y_{it}|L_{it}^{Real}, L_{it}^{All}, a_{it}, t, i] = f(L_{it}^{Real}) + f(L_{it}^{All}) + g(a_{it}) + \gamma(t) + \alpha_i$

We then estimate the model with a full set of fixed effects by OLS:[24]

$Y_{it} = \sum_{k=-9}^{9} \beta_k^{Real} 1_{\{L_{it}^{Real}=k\}} + \sum_{k=-9}^{9} \beta_k^{All} 1_{\{L_{it}^{All}=k\}} + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} +$

[22]We drop observations where k is below -9 or above +9 because there are too few observations far away from death and the coefficients on these leads and lags are therefore imprecisely estimated. Results are qualitatively similar when all observations are kept.

[23]The results are qualitatively similar when interacting age and year fixed effects.

[24]We exclude observations with inventors below the age of 25 or above the age of 70 from the sample to reduce variance, but the results are similar when these observations are included. When the dependent variable is citation or patent counts, we use a Poisson estimator, with QMLE standard errors clustered at the deceased-inventor level. The Poisson estimator with individual fixed effects fails to converge in our sample, therefore we report results without individual fixed effects and, as a robustness check, we run the same specifications with a negative binomial estimator with fixed effects.

$\alpha_i + \epsilon_{it}$

The main difference between our specification and the specifications used in the existing literature relying on premature deaths for identification is that we include a set of leads and lags around death that is common to both real and placebo survivors (L_{it}^{All}), in addition to the set of leads and lags around co-inventor death for the real survivors (L_{it}^{Real}). This application of the standard difference-in-differences estimator[25] to our setting addresses the concern that age, year and individual fixed effects may not fully account for trends in lifetime earnings and patents around co-inventor death. An inventor must necessarily have invented a patent before the year of (real or placebo) co-inventor death and is more likely to have been employed at that time, even conditional on a large set of fixed effects. Therefore, the construction of the sample of survivor inventors might mechanically induce a bias that the fixed effects do not fully address, and indeed we find that the set of leads and lags L_{it}^{All} has substantial predictive power for certain outcomes like employment. Intuitively, the leads and lags that are common to both real and placebo survivors (L_{it}^{All}) capture the mechanical effects, while the leads and lags that are specific to the real survivors (L_{it}^{Real}) capture the causal effect of co-inventor death.

Formally, if $E[1_{\{L_{it}^{All}=k\}}\epsilon_{it}|L_{it}^{Real}, L_{it}^{All}, a_{it}, t, i] = 0 \ \forall (t,k)$, then $\beta^{Real}(k)$ gives the causal effect of co-inventor death on the outcome of interest k years after death. Appendix D formally derives what is identified in this model and how the predictive effects $\{\beta^{Real}(k)\}_{k=-9}^{9}$ can be used to probe the validity of the research design and identify causal effects. It also compares our specification to those commonly used in the literature using premature deaths for identification.

In the next subsection, we use specification (1) to confirm the validity of the research design

[25]In the standard difference-in-difference estimator, treatment occurs at only one point in time and the regression includes an $After$ dummy and a $After \times Post$ dummy. In our setting, where co-inventors death are scattered over time, L_{it}^{All} plays a role analogous to the $After$ dummy and L_{it}^{Real} plays a role analogous to the $After \times Post$ dummy.

and study the dynamics of the effect. To summarize the results and discuss magnitudes, we employ a second specification, with a dummy turning to one after the time of co-inventor death for real survivor inventors ($AfterDeath_{it}^{Real}$) and another dummy turning to one after the time of co-inventor death for both real and placebo survivor inventors ($AfterDeath_{it}^{All}$). Under our identification assumption, β^{Real} gives the average causal effect of death.[26] This specification is as follows:

$$Y_{it} = \beta^{Real} AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it}$$

2.3.3. Results

Figure 3 reports the point estimates and 95% confidence interval for the coefficients β_k^{Real}, obtained from specification (1). Four outcome variables are considered: total earnings, labor earnings, non-labor earnings and citations. The point estimate on the lag turning to one in the year preceding death is normalized to 0 and inference is carried out relative to this lag.[27] We observe no pre-trending for any of the outcome variables, which lends credibility to the research design. The effect of co-inventor death on compensation and patents appears to manifest itself gradually: total earnings, labor earnings, non-labor earnings and citations all start to decline gradually after the death of a co-inventor. In line with the event studies in Figure 2, the nonparametric fixed effects for each lead and lag around death thus indicate that the nature of the effect is a change in the slope of the outcomes, rather than a level shift, and that co-inventor death has effects beyond short-term disruption of teamwork. As further discussed in Section IV, the gradual nature of the effect is consistent with the view that co-

[26] We have relatively more deaths occurring later in our sample and, as a result, β^{Real} gives more weight to the causal effects of death in the short-run after death and less weight to long-run effects. All results in the paper are about the average treatment effect on the treated.

[27] The full set of leads and lags L_{it}^{Real} always sum up to one for the survivor inventors and our specification includes individual fixed effects, therefore one of the leads and lags must be "normalized" to one. Appendix D discusses this standard normalization more formally.

inventor death impedes future co-invention activities: innovation is a stochastic process and the placebo survivors gradually outperform the real survivors.[28]

The magnitude of the effects is large. Eight years after the time of co-inventor death, the real survivor inventors' total earnings are $7,000 lower (4% of mean total earnings in the sample of survivors), their labor earnings are about $5,800 lower (3.8% of mean labor earnings in the sample of survivors) and their citation-weighted patent production is 15% lower than it would have been had they not experienced the premature death of a co-inventor.[29] About 80% of the total decline in earnings is due to a decline in labor earnings. We formally test the hypotheses that the point estimates are all the same before and after co-inventor death with a F-test, reported in Appendix Table B1 - we can never reject that the point estimates are all similar before death, but we can after death.

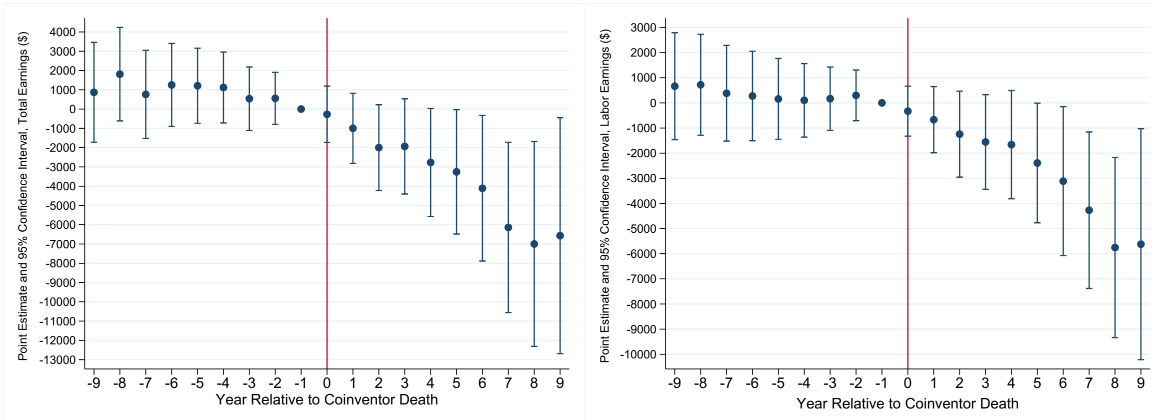
In order to reduce noise, we use specification (2), with a single indicator turning to one after the year of co-inventor death for real survivor inventors. The results are reported in Table 2. We use thresholds corresponding to the extensive margin, the 25th, 50th and 75th percentiles of the total earnings and labor earnings distributions to characterize heterogeneity in the effect across the income distribution.

[28] Bell et al. (2015) conduct event studies of inventor labor and non-labor earnings around the time of patent application and find that inventors' returns to innovation materialize gradually around the time of patent application in the form of both labor and non-labor earnings.

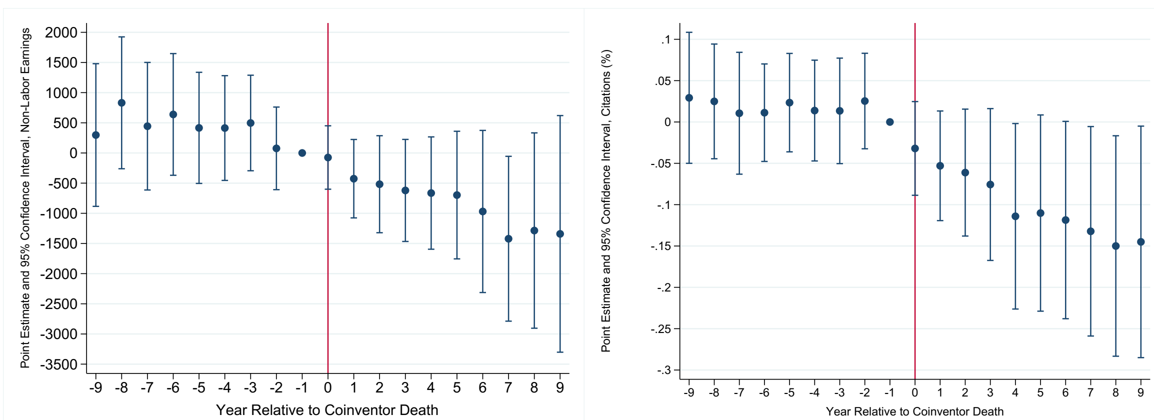
[29] The magnitude of the decline in citation-weighted patents is in line with the literature on peer effects in science. In life sciences, Azoulay et al. (2010) find that collaborators experience a 8% decline in quality-adjusted publications after the death of a "star." Oettl (2012) finds a corresponding decline of 16% in immunology. Based on the dismissal of Jewish scientists by the Nazi government, Waldinger (2012) shows that losing a coauthor of average quality reduces the average researcher's productivity by 13% in physics and 16.5% in chemistry.

Figure 2.3.: Dynamic Causal Effects of Co-inventor Death

Panels A and B: Survivor Inventor's Total Earnings and Labor Earnings



Panels C and D: Survivor Inventor's Non-Labor Earnings and Adjusted Forward Citations Received



Notes: Panels A to D of this figure shows the estimated β_k^{Real} coefficients from specification (1) for four outcome variables. Standard errors are clustered around the deceased inventors. Under the identification assumption described in Section III.B, β_k^{Real} gives the causal effect of co-inventor death in year k relative to co-inventor death. In panel D, the outcome variable is the count of forward citations received on patents the survivor applied for in a given year. Therefore, this variable reflects the timing and quality of patent applications by the survivor, not the timing of citations. Adjusted forward citations are winsorized at the 0.1% level. Dollar amounts are reported in 2012 dollars. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Appendix Table B2 shows that the results are similar on a balanced panel. For more details on the outcome variables, refer to Section II.C.

Table 2.2.: Causal Effects of Co-inventor Death

Panel A: Survivor Inventor's Total Earnings and Non-Labor Earnings

	Total Earnings	>p25	>p50	>p75	Non-Labor Earnings
$AfterDeath^{Real}$	-3,873***	-0.01531***	-0.0107**	-0.00772**	-1,199**
s.e.	(910)	(0.00434)	(0.00457)	(0.0039)	(498)
$AfterDeath^{All}$	- 223	0.00036	0.00066	-0.00068	651*
s.e.	(537)	(0.00285)	(0.00314)	(0.00297)	(378)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	OLS	OLS

Panels B and C: Survivor Inventor's Labor Earnings and Patents

	Labor Earnings	>0	Patent Count	Citation Count
$AfterDeath^{Real}$	-2,715***	-0.00913***	-0.09121***	-0.09024***
s.e.	(706)	(0.00315)	(0.02063)	(0.02326)
$AfterDeath^{All}$	-38	-0.0051**	0.00055	0.04084
s.e.	(480)	(0.00221)	(0.01776)	(0.03016)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	No	No
# Observations	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2). Column 1 reports the results for labor earnings. In column 2, the outcome variable is an indicator equal to one when the inventor receives a W-2, i.e. has positive labor earnings. The outcome variables for columns 3 to 5 are indicator variables equal to one when the inventor's labor earnings are above the specified quantile of the labor earnings distribution. The dollar value of these quantiles is reported in Table 1. Under the identification assumption described in Section III.B, β^{Real} gives the causal effect of co-inventor death on these various outcomes. Appendix Table B2 shows that the results are similar on a balanced panel. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 2 shows large and statistically significant coefficients β^{Real} for all outcome variables, consistent with the dynamic specifications reported in Figure 3. The effect exists across the distribution of adjusted gross income, and it seems larger in lower quantiles - a finding we will probe further in Section IV. Interestingly, β^{All} is significant for two outcomes: non-labor earnings and the extensive margin of labor earnings. The point estimates are large in magnitude relative to the point estimates for β^{Real}, which shows that controlling for mechanical patterns is important to avoid bias, even when age, year and individual fixed effects are included. Panel C of Table 2 shows that co-inventor death has large and significant effects for both the quantity of quality of patents produced by survivor inventors.[30]

2.3.4. Robustness Checks

Balanced Panel. We have confirmed that our results are robust to restricting attention to a balanced panel, focusing on survivors whose associated deceased passed away between 2003 and 2008 and considering a four-year window around death for each of these survivors. The results are presented in Appendix Table B2 and are similar to the results using the unbalanced panel.

Dynamics. The finding that co-inventor death has a long-lasting effect is one of the most striking results of this paper. Appendix Table B3 confirms that the effect becomes larger over time in a statistically significant way, using a specification with an indicator turning to one for observations more than four years after death (which reduces the noise reflected by the standard errors shown on Figure 2). A potential concern when studying the dynamics of the effect is related to how unbalanced the panel is with respect to years before and after the death of the co-inventor. For example, recent deaths have many pre-death observations but few post-death observations while the opposite holds for early deaths in the sample.

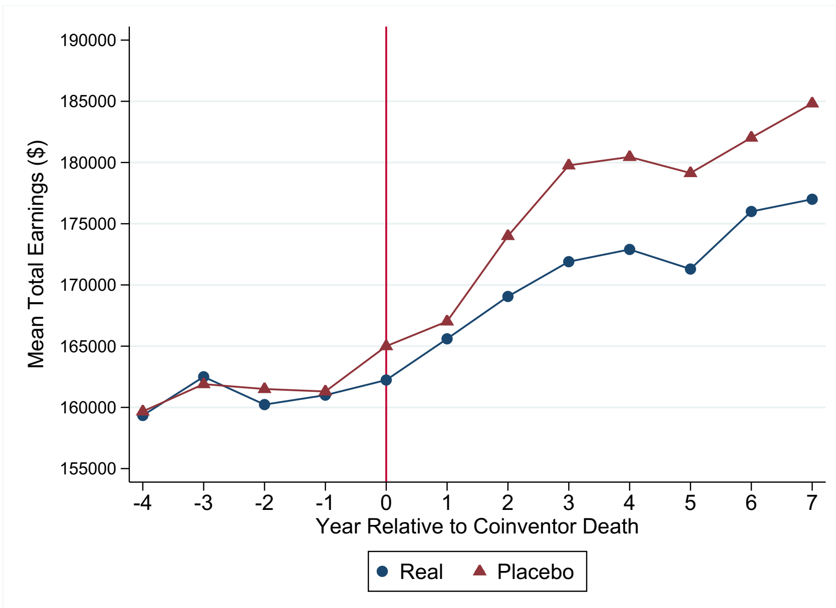
[30] The results for β^{Real} reported in Table 2 are the same when running the following specification, which replaces $AfterDeath_{it}^{All}$ in specification (2) with a full set of leads and lags around death (L_{it}^{All}):

$$Y_{it} = \beta^{Real} AfterDeath_{it}^{Real} + \sum_{k=-9}^{9} \beta_k^{All} 1_{\{L_{it}^{All}=k\}} + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it}$$

We have also checked that the results obtained with the Poisson estimator for count data are qualitatively similar when using OLS instead.

The dynamic specification can confound true dynamics due to the changing composition of the sample.[31] To address this issue, Figure 4 shows the path of total earnings for real and placebo survivor inventors experiencing death of their co-inventor between 2003 and 2005. This allows us to track the same individuals over time and confirms that the effect of coauthor death is indeed gradual and long-lasting. The regression results are presented in Appendix Table B4 and are qualitatively similar to the findings reported in Figure 3.

Figure 2.4.: Path of Total Earnings for Survivors with Co-inventor Death in 2003-2005



Notes: This figure shows the path of mean total earnings for real and placebo survivor inventors around the year of co-inventor death. The sample is restricted to the 4,812 co-inventors of the 1,764 real and placebo deceased with a year of death between 2003 and 2005. Inventor-year observations are dropped if the lag relative to co-inventor death is greater than seven years or if the lead relative to death is greater than four years. The panel is balanced: we observe the same inventors over a period of twelve years. Appendix Table B4 reports the results of the regression analysis in this sample.

[31] For example, it could be that inventors who experience death of a coauthor earlier in the sample are of higher ability than inventors who experience death of a coauthor later in the sample, which would manifest itself as larger long-run than short-run effects of death that are entirely due to changing sample composition rather than dynamic cumulative impacts. Similarly, one could imagine that earlier deaths in the sample had a bigger impact than later deaths but the impacts are constant following death: again, this would induce larger long-run than short-run effects, resulting from changing composition rather than dynamic cumulative impacts.

Anticipation. Another potential concern with our design is that co-inventor death may result from a lingering health condition. To investigate this hypothesis, we study tax deductions for high medical expenditures claimed by the deceased on their personal income tax return.[32] As shown in Appendix Figure B1, we find that seventy-five percent of deceased inventors do not claim any such deduction, but twenty-five percent claim a deduction in the year preceding death as well as in the year of death, and a small number claim deductions starting several years before death. As a robustness check, we repeat our analysis by excluding survivors whose associated deceased had a positive amount of tax deductions for high medical expenses in any year before death. We find that our results strengthen, as shown in Appendix Table B5. The point estimates for the various outcomes increase by about 10% (in absolute value). Intuitively, when the co-inventor is impaired before the time of death, our estimate of the causal effect on the survivors is biased downward because part of the effect starts before the time of death. This robustness check confirms that anticipation effects result in a downward bias and shows that the magnitude of the bias is relatively small.

Matching Strategy. We have investigated an alternative matching strategy, identifying a control group of placebo survivor inventors using propensity score reweighting, after estimating the propensity score on total earnings, labor earnings, year of birth and patent applications of the deceased inventors in the years preceding death. The results with this empirical strategy are reported in Appendix Figure B2 and Appendix Table B6 and are similar to the results using the real and placebo deceased exact match strategy.

Citations. Appendix Table B7 reports the causal effect of co-inventor death on a series of alternative measures of citations. Specifically, we consider in turns measures of citations that count only citations received in 3-year or 5-year citation windows after the time of grant or application (in order to address censoring), and that take into account only applicant-added or examiner-added citations. We find a large and statistically-significant effects, with

[32]This information is available on IRS form 1040.

magnitudes similar to Table 2. Appendix Table B8 shows the robustness of the citation results using a negative binomial estimator with individual fixed effects instead of a Poisson estimator.

Technology Classes. We check that our results are consistent across technology classes. Appendix Table B9 shows that, for the various outcome variables of interest, the effect of co-inventor death is not significantly different across technology classes. Our results are therefore not driven by a particular technology class.

Inference Taking into Account the Match Step. Lastly, we implement the coupled bootstrap procedure presented in Abadie and Spiess (2015) so that our standard errors reflect the matching step. The results are robust, with slightly smaller standard errors as shown in Appendix Table B10.

2.4. Distinguishing Between Mechanisms

In this section, we show that the gradual decline in earnings and citations caused by the premature death of a co-inventor stems from the fact that the survivor lost a co-inventor with whom they were collaborating extensively. We first rule out alternative mechanisms that are not specific to the team, establishing that the effect does not result from the disruption of the firm or from diffuse network effects. Second, we show that the effect is not driven by asymmetric top-down spillovers from unusually high-achieving deceased inventors. Third, we demonstrate that the intensity of the collaboration between the deceased and the survivor inventors prior to death is an important predictor of the magnitude of the effect. Fourth, we document that the majority of the effect results from the fact that the survivor can no longer co-invent with the deceased. Indeed, when considering only patents that were invented by the survivor without the deceased, the effect becomes much smaller. We also show that team-specific capital spans firm and geographic boundaries. Finally, we discuss other possible mechanisms consistent with the evidence.

2.4.1. Firm Disruption and Network Effects Are Not the Primary Mechanism

To test whether the effect documented in Section III results from the disruption of the firm or from diffuse network effects, we consider the groups of real and placebo coworkers and second-degree connections.[33] Figure 5 shows that the real and placebo coworkers and the real and placebo second-degree connections follow similar earnings paths both before and after the year of death of their associated deceased.[34] Appendix Figure C1 shows similar results for the paths of labor earnings and citations. This stands in sharp contrast with the diverging paths of real and placebo survivors after co-inventor death, as presented in Figure 3.

Table 3 reports the results obtained from specification (2) and shows that the premature death of an inventor has no significant negative effect on their coworkers and second-degree connections. The point estimates for the various outcome variables are generally one or two orders of magnitude smaller than the point estimates obtained for the direct co-inventors and are relatively precisely estimated.

For the coworkers, we find small and significant *positive* effects of an inventor's death on their coworkers' likelihood of being employed as well as on their patent and citation counts. Therefore, the large negative effect on the direct co-inventors of the deceased documented in

[33]The coworkers are the inventors who were in the same firm as the deceased in the year prior to death. The second-degree connection are the co-inventors of the co-inventors of the deceased. Refer to Section II for more details about the definition of these groups and the construction of the sample.

[34]The path of earnings for coworkers and second-degree connections - whether real or placebo - exhibits strong curvature around the time of (real or placebo) death. This curvature is partly captured by year and age effects. It also results from the fact that we impose that the coworkers should be employed in the year preceding death and that the second-degree connection should have co-invented with the survivors prior to death.

Section III do not result from the disruption of the firm or the R&D lab following an inventor's death.[35] The positive effect on coworkers may result from substitutability between inventors at the same firm: an inventor's earnings and patent production might rise after the death of a coworker because it increases this inventor's chance of being promoted and their access to resources within the firm.[36]

For the second-degree connections, we find no statistically significant effect on any of the outcomes. The point estimates are close to zero and we can reject at the 5% confidence level any effect of a magnitude larger than one half of the effect documented for the direct co-inventors. This evidence provides a test of competing models of strategic interactions in networks. If the dominant force is a substitution effect as in Jackson and Wolinsky (1996), then we should find that the second-degree connections benefit from the death. But if strategic complementarities dominate as in Bramoulle et al. (2014), then the death should negatively affect the second-degree connections. Our finding that, on net, the effect on second-degree connections is negligible means that network effects are not first-order, as opposed to the direct impact on co-inventors.

Therefore, we can rule out firm disruption and network effects as primary mechanisms explaining the effect documented in Section III.[37] Moreover, the analysis of the effect on coworkers and second-degree connections generated new insights about the innovation production function: the results suggest that inventors within a firm are substitutable while there is no strong complementarity or substitutability patterns between inventors who are two nodes

[35] We provide additional evidence confirming this fact by showing that the effect persists for co-inventors located in different firms at the time of death (Table 9) and that the magnitude of the effect is not correlated with firm size (Appendix Table C6).

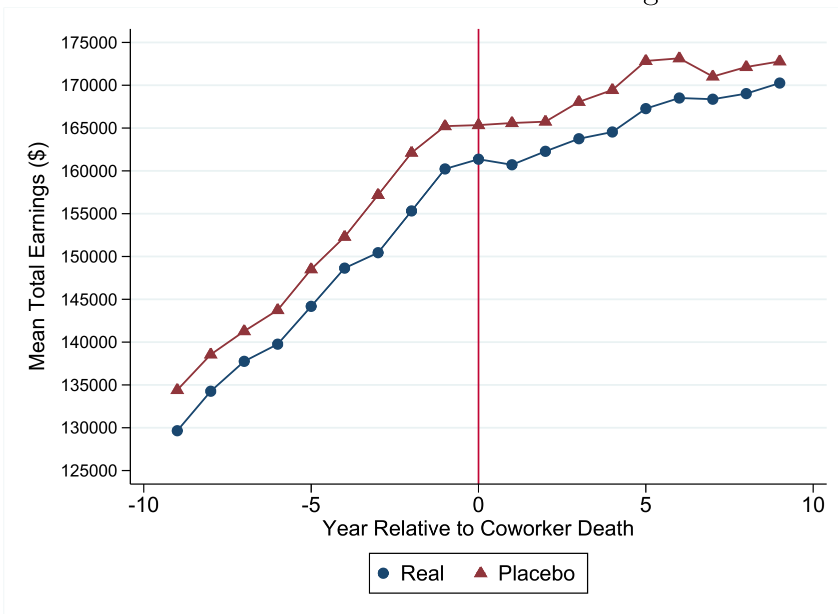
[36] Further exploration of the mechanism at play for coworkers is beyond the scope of this paper, but our results are consistent with those obtained in parallel work by Jaeger (2015), who studies small firms in Germany rather than the population of inventors, as we do.

[37] We have also constructed a "citation network" of inventors who cited the deceased before their death but who were not among their direct co-inventors, second-degree connections or coworkers. We do not find evidence of statistically significant negative effects. These results are not surprising, given how diffuse citation networks are, but they establish that the effect is not driven by linkages in idea space. These results are available from the authors upon request.

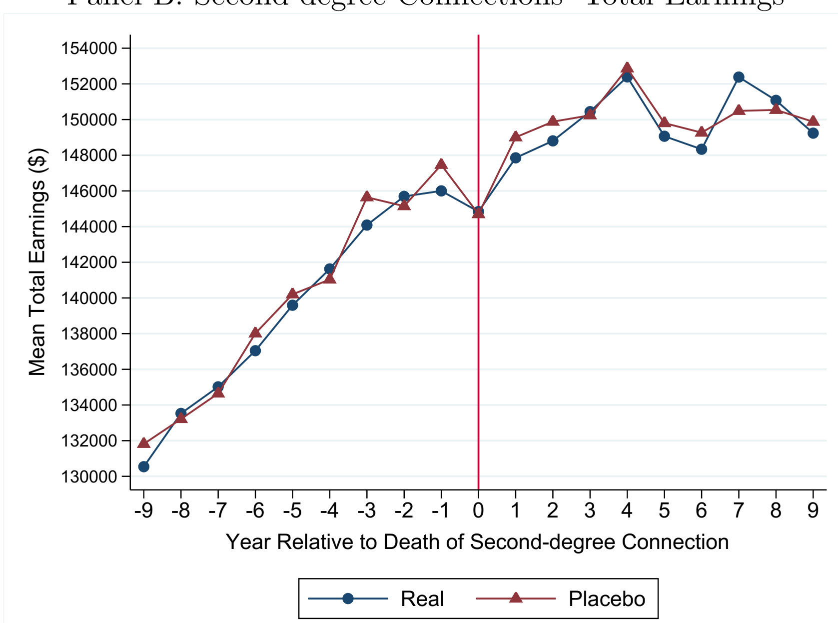
away in the co-invention network.

Figure 2.5.: Path of Outcomes for Coworkers and Second-Degree Connections Around Year of Death

Panel A: Coworkers' Total Earnings



Panel B: Second-degree Connections' Total Earnings



Notes: This figure shows the path of mean total earnings for real and placebo coworkers as well as for real and placebo second-degree connections around the year of death of their associated deceased. The sample includes all real and placebo inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Refer to section II.B for more details on the sample and to section II.C for more details on the outcome variables.

Table 2.3.: Causal Effects of Inventor Death on Coworkers and Second-degree Connections

Panel A: Effect on Coworkers

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	207	236	0.00639**	0.0249*	0.0148**
s.e.	(571)	(582)	(0.00296)	(0.0131)	(0.00713)
$AfterDeath^{All}$	-745	-682	-0.00536**	-0.0366**	-0.00976**
s.e.	(818)	(853)	(0.00215)	(0.01664)	(0.00416)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	335,708	335,708	335,708	335,708	335,708
# Coworkers	28,192	28,192	28,192	28,192	28,192
# Deceased	3,988	3,988	3,988	3,988	3,988
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and β^{All} from specification (2) in the sample of coworkers. The five outcome variables are as follows: (1) total earnings; (2) labor earnings; (3) an indicator equal to one when the inventor receives a W-2, i.e. has positive labor earnings; (4) the number of patents the coworker applied for in a given year; (5) the number of forward citations received on patents that the coworker applied for in a given year (therefore, this variable reflects the timing and quality of patent applications by the survivor, not the timing of citations). Under the identification assumption described in Section III.B, β^{Real} gives the causal effect of coworker death on these various outcomes. Inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Appendix Table C1 shows that the results are similar on coworker sample keeping firms of all sizes. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel B: Effect on Second-degree Connections

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-159	-9	0.0027	-0.00258	-0.02346
s.e.	(548)	(506)	(0.00325)	(0.02115)	(0.0210)
$AfterDeath^{All}$	-618	-684	-0.00618*	-0.08121**	-0.0208
s.e.	(749)	(565)	(0.00367)	(0.0363)	(0.02625)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	265,421	265,421	265,421	265,421	265,421
# Second-degree Connections	23,331	23,331	23,331	23,331	23,331
# Deceased	4,183	4,183	4,183	4,183	4,183

Notes: This panel reports the estimated coefficients β^{Real} and β^{All} from specification (2) in the sample of second-degree connections. The five outcome variables are as in Panel A. Inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.4.2. Top-Down Spillovers Are Not the Driving Force

As documented in Section II, some teams are composed of inventors of similar age and compensation levels, while in others there are large gaps in age and compensation levels between team members. We study whether these patterns are important predictors of the heterogeneity in the average effects documented in Section III. In particular, we want to test whether the effect is driven by the death of "superstar" inventors or, more generally, by inventors of higher ability level than their associated survivors.

To do so, we repeat the estimation of the coefficient of interest, β^{Real}, by using specification (2) in different subsamples of the data. We partition the data depending on the quartile in which the total earnings of the (real and placebo) deceased and the (real and placebo) survivor inventors fall three years before the year of (real and placebo) death. The sample sizes in each subsample are given in Appendix Table C2. This way of inferring relative ability levels can potentially create mean reversion patterns. For instance, it could be that survivor inventors who are in the first quartile of the earnings distribution three years before co-inventor death suffered from temporary shocks and that their earnings tend, on average, to increase afterwards. The use of our control group of placebo survivor inventors is sufficient to alleviate these concerns if the income processes are similar for the real and placebo survivor inventors prior to the death of the co-inventor (i.e. both groups are affected by mean reversion and other such patterns in similar ways). To investigate whether this is true, we examine the distribution of changes in total earnings for the years before the death of the co-inventor. The difference in this analysis relative to our earlier analysis in Section III is that we now want to ensure that the placebo survivor inventors are an appropriate control group for the distribution of changes in potential outcomes over time, not just for their mean. Table 4 shows that the distribution of earnings changes is very similar for the real and placebo survivor inventors.[38]

[38] We obtain similar results when considering changes of total earnings in levels as well as level or log changes

Table 2.4.: Distribution of Annual Changes in Log Total Earnings Before Co-inventor Death

		Mean	SD	10pc	25pc	50pc	75pc	90pc
Total Earnings	Real	0.039	0.457	-0.0026	0.0169	0.035	0.0867	0.1436
	Placebo	0.040	0.461	-0.0024	0.0188	0.036	0.0844	0.1401

Notes: This table reports the distribution of year-to-year changes in log total earnings for real and placebo survivor inventors before the year of co-inventor death. The distributions are very similar across the two groups, suggesting that the income processes are similar for both groups and that the placebo inventors can be used as a control group for the analysis reported in Table 5. The results are similar when considering annual changes in the level of total earnings, the log of labor earnings and the level of labor earnings. For more details on the sample, see Section II.B.

Table 5 reports the results of this analysis. Consider for instance panel A on total earnings (the findings are similar for panel C, on labor earnings). This panel shows three main findings. First, the effect is significant and large in magnitude when the deceased and the survivor are in the same earnings quartile, i.e. are of similar seniority levels. This rejects the hypothesis that the effect documented in Section III is entirely driven by top-down spillovers from "superstar" inventors, because the effect persists for inventors of similar seniority levels. Second, holding constant the earnings quartile of the survivor, the effect is increasing in the earnings quartile of the deceased, showing that co-inventors of a higher seniority level are more difficult to substitute for. Third, the effect is not significant when the deceased is in a lower earnings quartile than the survivor. Although the point estimates are imprecisely estimated, it suggests that co-inventors of a lower seniority level are not a source of specific value for an inventor. The fact that lower ability team members suffer from the loss of higher ability team members, while in contrast higher ability team members are largely unaffected by the loss of a lower ability peer, could indicate that lower ability inventors extract "rents" from their collaboration with high ability co-inventors. However, this "rent" hypothesis cannot explain the large effect we find for team members of similar ability levels.

Moreover, panels C and D of Table 5 show that mechanical patterns (due to mean reversion or other statistical effects) play a very important role. These panels show that there are

for labor earnings.

strong mean-reversion patterns: survivors in the lowest earnings quartile before (placebo) co-inventor death tend to perform better after the year of death, while survivors in the highest earnings quartile before (placebo) co-inventor death tend to perform worse after the year of death. Therefore, year, age and individual fixed effects are not sufficient to account for trends in earnings around the time of co-inventor death and it is important to include the $AfterDeath^{All}$ dummy introduced in specification (2).

We have conducted a series of robustness checks of these results. First, instead of running the analysis in different subsamples as in Table 5, we ran regressions with a linear interaction of the $AfterDeath$ indicator with the quartile difference or the level difference in the labor earnings levels of the survivor and the deceased, as well as with the age difference between the survivor and the deceased. Second, we have checked that the results are similar with other metrics of relative ability levels, namely the relative level of total earnings and the relative citation levels three years before death.[39] We find that the causal effect is larger when the survivor inventor is of lower ability or seniority than the deceased and the effect is still significant for inventors of equal ability or seniority levels.

[39] A limitation of using relative citations before death is that the survivor and the deceased have often co-invented most of their patents together, therefore relative earnings appear to be a better signal of relative seniority.

Table 2.5.: Heterogeneity in Effect by Relative Ability Levels of Co-inventors

Panel A: Heterogeneity in the Causal Effect of Co-Inventor Death on Total Earnings

Deceased / Survivor Earnings Quartile	1	2	3	4
1	-2,652*	-1,301	1,298	902
s.e.	(1,553)	(1,328)	(1,680)	(1,081)
2	-3,573*	-2,798**	-810	-1,308
s.e.	(2,111)	(1,178)	(1,675)	(1,278)
3	-5,656**	-4,151**	-3,243**	-2,939
s.e.	(2,612)	(1,968)	(1,632)	(2,562)
4	-6,566*	-5,132**	-4,853*	-7,037**
s.e.	(3,450)	(2,530)	(2,650)	(3,256)

Notes: This panel reports the estimated coefficient β^{Real} from specification (2), with total earnings of the survivors as the outcome variable, in sixteen subsamples of the data. Each of these subsamples corresponds to a different combination of the total earnings quartiles of the survivor and the deceased. The earnings quartiles are computed three years before death and sample sizes for each subsample are given in Appendix Table C2. Under the identification assumption described in Section III.B, β^{Real} gives the causal effect of co-inventor death on total earnings. For instance, the panel shows that if the survivor and the deceased were both in the lowest quartile of total earnings three years before death, the causal effect of co-inventor death on the survivor was a decline of $2,652 in total earnings. Amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. $*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Panel B: Mean Reversion Patterns in Total Earnings Around Co-inventor Death

Deceased / Survivor Earnings Quartile	1	2	3	4
1	14,763***	3,373	-1,397	-18,977***
s.e.	(2,138)	(2,136)	(2,844)	(3,994)
2	14,493***	380	1,536	-13,665***
s.e.	(2,329)	(1,356)	(1,845)	(2,947)
3	15,237***	3,410**	1,087	-18,473***
s.e.	(2,401)	(1,425)	(2,200)	(3,803)
4	17,183***	-671	3,384	-13,539***
s.e.	(4,243)	(2,681)	(2,599)	(3,814)

Notes: This panel reports the estimated coefficient β^{All} from specification (2), with total earnings of the survivors as the outcome variable, in sixteen subsamples of the data. Each of these subsamples corresponds to a different combination of the total earnings quartiles of the survivor and the deceased. The earnings quartiles are computed three years before death and sample sizes for each subsample are given in Appendix Table C2. β^{All} gives the predictive effect of placebo co-inventor death on total earnings, conditional on year, age and individual fixed effects. For instance, the panel shows that if the placebo survivor and deceased were both in the lowest quartile of total earnings three years before death, then after the placebo death of their co-inventor, the total earnings of placebo survivor inventors tended to increase by $14,763. Amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. $*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.4.3. The Effect Is Driven by Close-Knit Teams

We consider various measures of collaboration intensity between deceased and survivor inventors, which Table 6 shows vary widely in our sample. Specifically, we use the number and share of patents the survivor inventor co-invented with the deceased, collaboration length (defined as the number of years between the first and last joint patent application between the survivor and the deceased), and collaboration recency (defined as the numbers of years between the death of the co-inventor and the application for the last co-invented patent with the survivor).

Table 2.6.: Collaboration Patterns Between Deceased and Survivor Inventors Before Death

Variable	Sample	Mean	SD	10pc	25pc	50pc	75pc	90pc
# Patents	Real	8.114	17.285	1	1	3	9	18
	Placebo	7.41082	12.757	1	1	3	8	18
# Co-patents	Real	1.702	1.502	1	1	1	2	3
	Placebo	1.6108	1.394	1	1	1	2	3
Co-patent Share	Real	54.61	37.75	7.692	18.75	50	100	100
	Placebo	54.55	37.81	8.33	18.18	50	100	100
Collaboration Length	Real	0.8208	1.7393	0	0	0	1	3
	Placebo	0.7593	1.7050	0	0	0	1	3
Collaboration Recency	Real	6.1125	3.9756	1	3	6	9	12
	Placebo	5.673	4.0078	1	2	5	8	12
# Real Survivors	14,150							
# Placebo Survivors	13,350							

Notes: This table documents the heterogeneity in the intensity of collaboration between the deceased and survivor inventors in the years before (real or placebo) death. The variables are defined as follows: (1) # patents is the number of patents of the survivor before co-inventor death; (2) # co-patents is the number of patents co-invented by the survivor and the deceased before co-inventor death; (3) co-patent share is the share of the survivor's patents that were co-invented with the deceased before death; (4) collaboration length is the number of years that elapsed between the first and last joint patent application between the survivor and the deceased; (5) collaboration recency is the number of years that elapsed between the application year for the last patent co-invented by the survivor and the deceased and the year of co-inventor death. For more details on the sample, refer to Section II.B.

To examine whether heterogeneity in collaboration strength predicts heterogeneity in the causal effects, we set up the following specification:

$$Y_{it} = \begin{array}{l} \beta^{Real} AfterDeath_{it}^{Real} + \eta^{Real} X_i \cdot AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} + \\ \eta^{All} X_i \cdot AfterDeath_{it}^{All} + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array} \quad (2.1)$$

where X_i is a vector including all variables listed in Table 6, as well as the age of the survivor inventor at the time of death. The vector X_i is demeaned so that the point estimates for β^{Real} and β^{All} are left unaffected.

Table 7 reports the results for the relevant interaction terms. It shows that the various proxies for the intensity of the collaboration between the survivor inventor and the deceased (co-patent share, collaboration length and collaboration recency) are strong predictors of the magnitude of the causal effect of co-inventor death on the various outcomes. The point estimates are all negative and statistically significant. Using the standard deviations reported in Table 6 for the various regressors and the magnitude of the causal effects reported in Table 2, we can gauge the magnitude of the predictive effects. A one standard deviation increase in the share of copatents explains 75% of the average effect on total earnings, 78% of the average effect on labor earnings, 70% of the average effect on patent count, and 54% of the average effect on citation count. Similarly, a one standard deviation increase in collaboration length explains 47% of the average effect on total earnings, 33% of the average effect on labor earnings, 46% of the average effect on patents, and 53% of the average effect on citations. Lastly, a one standard deviation increase in collaboration recency explains 45% of the average effect on total earnings, 52% of the average effect of labor earnings, 22% of the average effect on patents, and 21% of the average effect on citations. This indicates that the effect is driven by the loss of a co-inventor that the survivor was collaborating with extensively.[40]

[40] Our results differ markedly from Azoulay et al. (2010), who do not find collaboration intensity to be predictive of the magnitude of the effect of the death of a superstar on their coauthors. It could be due to the fact that top-down spillovers, which are not the driving force in our data, do not strongly depend on the intensity of collaboration. Azoulay et al. (2010) interpret their results as evidence for very diffuse spillovers in intellectual space for "star scientists." In contrast, our results provide evidence for very circumscribed spillovers in collaboration space for the typical inventor.

Table 2.7.: Heterogeneity in Effect by Intensity of Collaboration Between Deceased and Survivor Inventors

η^{Real}	Total Earnings	Labor Earnings	Non-Labor Earnings	Patent Count	Citation Count
Co-patent Share	-75.132***	-56.669***	-17.236**	-0.00172**	-0.0013*
s.e.	(22.552)	(17.164)	(8.342)	(0.00085)	(0.00069)
Collaboration Length	-1,063.253***	-523.296**	-323.296***	-0.0245**	-0.02892*
s.e.	(405.382)	(228.55)	(118.516)	(0.01072)	(0.01537)
Collaboration Recency	447.921***	360.281***	110.728**	0.00508**	0.00482*
s.e.	(145.592)	(139.825)	(50.95)	(0.00256)	(0.00266)
# Co-patents	42.163	64.029	20.231	0.0015	0.00127
s.e.	(107.372)	(121.255)	(431.156)	(0.01962)	(0.0124)
# Patents	-49.129	5.022	-60.001	-0.00642**	-0.00442**
s.e.	(57.941)	(39.44)	(40.223)	(0.00287)	(0.00181)
Survivor's Age at Death	104.78*	40.961	50.899	- 0.00243**	-0.00323**
s.e.	(62.774)	(49.876)	(40.85)	(0.001073)	(0.00129)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This table reports the estimated coefficients in the vectorη^{Real} from specification (3). The outcome variables reported in the five columns are total earnings, labor earnings, an indicator turning to one if the inventor receives a W2, the number of patents the survivor inventor applied for in a given year, and the number of forward citations received on patents that the survivor applied for in a given year (therefore, this variable reflects the timing and quality of patent applications by the survivor, not the timing of citations). The regressors are defined in the main text as well as in Table 6 and are demeaned so that the point estimates for the average causal effects are identical to Table 2. Standard errors are clustered around the deceased inventors. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

2.4.4. Team-Specific Capital as a Likely Mechanism

Taken together, the evidence suggests that the gradual decline in earnings and citations following the premature death of a co-inventor results from the fact that the survivor lost a partner with whom they were collaborating intensely. The heterogeneity in the effect by intensity of collaboration, as well as the pervasive nature of the effect across various kinds of teams, makes team-specific capital a likely mechanism. The difficulty of building a similar relationship with another inventor may result from high search costs or from the fact that the quality of the relationship improved endogenously over time, in response to relationship-specific investments made by each of the co-inventors.

Consistent with the team-specific capital interpretation, we find that the effect of co-inventor death is much larger in the context of joint production and exists across firm and geographic boundaries. First, we repeat the analysis of the effect of co-inventor death on the patents of the survivor, but now we only consider patents that were not co-invented with the deceased.[41] Table 8 reports that, for the various measures of patent production and citations, we consistently find a significant and negative effect of co-inventor death. Continued interaction with a co-inventor therefore benefits an inventor beyond co-inventions, which is consistent with the view of teams as a vehicle for knowledge transmission. However, the magnitude of the effect on the survivor's patents outside of patents with the deceased is much smaller (around -3%) relative to the effect on the total number of patents of the survivor documented in Table 2 (around -9%). This suggests that the main value of team-specific capital comes in the form of co-inventions and that the effect results from the fact that the survivor can no longer engage in joint projects with the deceased.[42]

Second, we show that the effect persists for inventors located in different firms and in different commuting zones. Panel A of Table 9 shows that the effect of co-inventor death on labor earnings is entirely driven by survivors who were in the same firm as the deceased at the time of death. In contrast, the second column shows that the effect of co-inventor death on non-labor earnings is similar regardless of whether or not the survivor and the deceased were in the same firm.[43] Panel B of Table 9 shows a similar pattern based on the location of

[41] Note that legal requirements impose that all inventors should be listed on a patent, otherwise the patent could be invalidated in court. We can therefore be confident that the patents that do no list the name of the deceased were indeed invented without the active collaboration of the deceased.

[42] Note that our results are very different from Azoulay *et al.* (2010), who find that the death of a "star" scientist causes a decline of similar magnitude in scientific publications with and without the deceased. In our setting, the importance of joint production between the deceased and the survivor is consistent with the gradual effect documented in Section III: innovation is a stochastic process and the placebo survivors gradually outperform the real survivors.

[43] As has been documented in prior work (e.g. Crescenzi et al., 2015), co-inventors may be working in different

survivor and deceased inventors across commuting zones.[44] Therefore, team-specific capital is not tied to firm or geographic boundaries.

Table 2.8.: The Causal Effect of Co-inventor Death On the Survivor Beyond Joint Production

	Only Considering Patents that Were Not Co-invented With the Deceased			
	Patent Count	Citation Count	Count of Patents with No Citations	Count of Patents in Top 5% of Citations
$AfterDeath^{Real}$	-0.03088**	-0.03571**	-0.03288**	-0.0084*
s.e.	(0.01525)	(0.01815)	(0.01525)	(0.00478)
$AfterDeath^{All}$	0.1162**	0.08578	0.05763	0.0247
s.e.	(0.05319)	(0.12013)	(0.08136)	(0.02271)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	No	No	No	No
# Observations	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428
Estimator	Poisson	Poisson	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2). The four outcome variables are as follows: (1) patent count is the number of patents the survivor inventor applied for in a given year, excluding all patents co-invented with the deceased; (2) citation count is the number of forward citations received on patents that the survivor applied for in a given year, excluding all patents co-invented with the deceased; (3) the count of patents with no citations is the number of patents that the survivor inventor applied for in a given year and that have never been cited as of December 2012, excluding all patents co-invented with the deceased; (4) the count of patents in the top 5% of citations is the number of patents the survivor inventor applied for in a given year that were in the top 5% of the citation distribution, excluding all patents co-invented with the deceased. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is more than 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

entities in the context of university - private sector partnerships or joint ventures. Cross-firm collaborations are common in sectors like biotech. Our definition of "firm" is based on EINs, which cover universities and public sector institutions.

[44] These findings are consistent with the view that the effect of co-inventor death on earnings primarily comes from the fact that the survivor is no longer able to work with his co-inventor on joint inventions. Indeed, inventors who are co-inventors but who work for different firms may be collaborating on joint projects outside of their work as employees. If they are successful, these projects are likely to result in an increase in non-labor earnings rather than labor earnings.

Table 2.9.: The Causal Effect of Co-inventor Death across Firm and Geographic Boundaries

Panel A: Within and Across Firms

	Labor Earnings	Non-Labor Earnings	Patent Count	Citation Count
$AfterDeath^{Real}$	-113	-1,225**	-0.07071**	-0.07892**
s.e.	(964)	(583)	(0.03321)	(0.0353)
$AfterDeath^{Real} \cdot SameFirm$	-3,974***	122	-0.05928	-0.05123
s.e.	(1,465)	(983)	(0.06956)	(0.04326)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes
# Observations	260,807	260,807	260,807	260,807
# Survivors	21,972	21,972	21,972	21,972
# Deceased	7,589	7,589	7,589	7,589
Estimator	OLS	OLS	Poisson	Poisson

Panel B: Within and Across Commuting Zones

	Labor Earnings	Non-Labor Earnings	Patent Count	Citation Count
$AfterDeath^{Real}$	-182	-1,411**	-0.09393***	-0.1229***
s.e.	(529)	(563)	(0.02901)	(0.02856)
$AfterDeath^{Real} \cdot SameCZ$	-4,049***	534	0.00093	0.0209
s.e.	(1,350)	(610)	(0.05512)	(0.0212)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	No	No
# Observations	292,752	292,752	292,752	292,752
# Survivors	24,686	24,686	24,686	24,686
# Deceased	8,579	8,579	8,579	8,579
Estimator	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

$$Y_{it} = \begin{array}{c} \beta^{Real} AfterDeath^{Real}_{it} + \beta^{All} AfterDeath^{All}_{it} + \widetilde{\beta^{Real}} AfterDeath^{Real}_{it} \cdot SameCZ + \widetilde{\beta^{All}} AfterDeath^{All}_{it} \cdot SameCZ \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where $SameCZ$ is an indicator variable equal to one when the survivor and the deceased were in the same commuting zone during the three years that preceded death. $SameCZ$ is equal to 0 when the survivor and the deceased were in different commuting zones during the three years that preceded death. We exclude from the sample the survivor-deceased pairs that were not always in the same commuting zone or always in a different commuting zone during the three years prior to death. 10.24% of the survivors are thus excluded. $SameCZ$ is equal to 1 for 55% of survivors in the sample. See Table 2 for details about the outcome variables. Standard errors are clustered around the deceased inventors. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A number of mechanisms in which team-specific capital plays no role may be able to explain our results but appear unlikely. First, emotional distress following the loss of a co-inventor may result in a decline in productivity - however, for this mechanism to be consistent with the patterns we have documented, emotional distress would need to be long-lasting, it should be larger when losing a high-achieving peer and it should cause labor earnings to fall only for inventors who work in the same firm. Second, the effect of co-inventor death might be driven by disruption of current work - however, we find the effect to be long lasting and we also find an effect on the survivor inventor's patents beyond co-inventions with the deceased. Third, the effect could be driven by a change in physical inputs available to survivor inventors. For example, after the death of a prominent inventor, the R&D lab might close down, or the start up may fail - however, we find that the effect exists for inventors working in different firms, as well as for co-inventors of average ability, and we find no negative spillover effect on coworkers in the same firm as the deceased. Fourth, the effect may be driven by a lower ability inventor exploiting a rent from their collaboration with a higher ability deceased - however, the effect persists for co-inventors of equal ability levels and there is an effect beyond joint production, on the survivor's patents beyond co-inventions with the deceased.

Thus, our results show that team-specific capital is important in an inventor's career because it facilitates co-inventions and - to a lesser extent - knowledge transmission. We have conducted interviews with patent inventors to confirm that this mechanism is plausible.[45] Moreover, it is in line with the notion that playing a repeated game with team members helps curb moral hazard in joint production and information exchange (e.g. Stein, 2008). Appendix C documents other heterogeneity patterns in the effect of co-inventor death - by firm size, survivor's age, survivor's co-inventor network size and survivor's citizenship status - which are of descriptive interest but are not statistically significant for most outcomes. Appendix C also shows that co-inventor death does not have a strong impact on the probability that an inventor starts new collaborations or changes firms, except if the inventor was in a

[45] We spoke with fourteen inventors in small start-ups as well as large R&D labs in Silicon Valley. They pointed out the difficulty of building good collaborative relationships and emphasized the long-lasting nature of successful collaborations, which often continue to exist across firm boundaries.

small firm before their co-inventor's death.

2.5. Conclusion

In this paper, we have shown that team-specific capital is an important ingredient of the typical patent inventor's lifecycle earnings and productivity, much like firm-specific capital is crucial for the typical worker (Topel, 1991). Exploiting the premature deaths of 4,714 inventors in a difference-in-differences research design, we find that a co-inventor's premature death causes a large and long-lasting decline in an inventor's labor earnings (- 3.8% after 8 years), total earnings (- 4% after 8 years) and citation-weighted patents (- 15% after 8 years). We find that this effect exists for various kinds of teams and is not limited to top-down spillovers within the team, although the effect is larger when the survivor inventor is of lower ability than the deceased inventor. Consistent with the team-specific capital interpretation, the effect is larger for more closely-knit teams and primarily applies to co-invention activities with the deceased. The paper also provides estimates of the causal effect of an inventor's death on coworkers and second-degree connections. We find that an inventor's earnings and patents are not significantly adversely affected by the premature death of a coworker at the same firm who is not a co-inventor, nor by the premature death of an inventor two nodes away in the co-inventor network. This evidence indicates that inventors are not difficult to replace from the perspective of their coworkers and second-degree connections, which underscores the unique role of teams.

Identifying the magnitude and nature of spillover effects between inventors is central to innovation and tax policy design, because the impact of any policy may depend greatly not just on a given inventor's behavior but on a "multiplier effect" that affects the broader innovation process. In this paper, we have established empirically the relevance of team-specific capital, which generates a multiplier effect between co-inventors. This multiplier effect may cause a wedge between the private and social returns to the accumulation of team-specific capital.[46] Our evidence suggests that the social returns to improving the match technology between inventors and encouraging the accumulation of team-specific capital may be very large. For instance, high-skill immigration policy might have a crucial role to play by increasing the supply of inventors - thus potentially reducing the fixed cost of finding a good match and making it easier for an inventor to find substitutes for their close collaborators - or by offering visa extensions to successful inventors - thus preserving team-specific capital that was built during the course of successful collaborations. Without further evidence on the exact mechanism at play, however, the policy implications of our findings can only be tentative.

The evidence and methodology described in this paper point to several promising directions for future research. First, the parameters of a structural model of team-specific capital formation could be estimated by using the premature death shock, simulating the model with respect to such a shock and fitting moments in the data. Second, it would be useful to examine whether significant spillover effects exist in some subsamples of the more diffuse networks we have considered, given that these effects are more likely to be genuine externalities introducing a wedge between the private and social returns to knowledge production. Finally, given the prevalence of teamwork in modern economies, investigating the role of team-specific capital in sectors of the economy beyond innovation and patents would be of great interest.

[46] Note that, on its own, our natural experiment cannot be used to conclude whether or not such a wedge exists. Perhaps the employer internalizes all effects, or perhaps the mobility of inventors across both teams and firms creates a wedge between private and social returns.

3. Patent Trolls and the Patent Examination Process[1]

3.1. Introduction

This paper sheds new light on the patent acquisition behavior of non-practicing entities (NPEs). NPEs, also known as patent assertion entities or patent trolls, garner profits exclusively from IP litigation and licensing without producing or selling goods. They have grown in prominence in the IP system over the past decade, attracting a large amount of scrutiny and debate along the way. In the words of President Obama, "[NPEs] don't actually produce anything themselves. They're just trying to hijack somebody else's idea and see if they can extort some money out of them."[2] The general unpopularity of NPEs within Congress has led to several attempts to pass legislation to curb their activity, including the Strong Patent Act[3] and the Innovation Act[4], but arguments over the definition of NPEs and the nature of their activities have prevented them from passing. In defense of their business model, NPEs argue that they serve as intermediaries that improve the efficiency of the market for ideas, by helping resource-constrained inventors and firms enforce their patents against infringing

[1]Co-authored with Josh Feng.
[2]Google Hangout Session, February 14th 2015.
[3]https://www.congress.gov/bill/114th-congress/senate-bill/632
[4]https://www.congress.gov/bill/113th-congress/house-bill/3309

entities.[5] By law, all issued patents are supposed to cover only "novel" and "non-obvious" inventions, but an average application gets less than 20 hours of patent examiner time, and a large proportion of the few patents later fully evaluated in court are held invalid. A blatantly invalid patent, which clearly would be overturned in court, would never be asserted and would thus cause no harm. The criticism of patent trolls is based on the concern that they puchase and assert patents that are not clearly invalid, but are "weak" - they may well be invalid, but nobody knows for sure without conclusive litigation.[6] On average, do NPEs tend to purchase "weak" patents? We give a positive answer to this question by documenting that, unlike regular firms, NPEs purchase and assert patents that were granted by a specific set of examiners, who tend to allow incremental patents with vaguely-worded claims that are therefore more likely to be "weak."

In addition to being the first large-scale empirical study of the patent acquisition behavior of patent trolls, this paper speaks to another important debate about the patent system: do patent examiners have a significant influence on the eventual outcomes and uses of the patents they grant, or do existing rules at the United States Patent Office (USPTO) constraint them such that they have little ability to reject of modify patent applications[7]? Answering this question is key to establish whether policy changes related to examination practices have an important role to play in reforming the Intellectual Property (IP) system. Recently, the United States Patent Office began implementing the so-called *Enhanced Patent Quality Initiative* (EPQI), which aims at improving patent correctness[8] and clarity[9] by train-

[5]As explained on Intellectual Ventures' website, "We purchase patents from individual inventors, start-ups, large corporations, research institutions, and everything in between. By acquiring these inventions, we provide capital to inventors and give their ideas a better chance of getting into the marketplace."

[6]As is well-known, most litigation cases involving patent trolls are held at the Eastern District Court of Texas, which is widely believed to have a pro-plaintiff bias.

[7]For instance, it may be very difficult for examiners to reject weak patents because the rules of the USPTO are such that the burden of proof of non-patentability rests on the examiners. This view is articulated in Lei and Wright (2009), who emphasize that other patent offices like the European Patent Office place much fewer constraints on the examiners.

[8]Correctness refers to correctly judging whether an application meets all patentability criteria.

[9]Clarity refers to whether the granted patent clearly defines the technology covered by it.

ing examiners and running pilots to detemine which examination practices can best achieve these two goals. In a recent blog post[10], Michelle Lee, Director of the USPTO, suggested that the renewed focus on clarity in the EPQI was necessary because of "the evolving patent landscape" and that improving patent clarity could reduce "needless high-cost court proceedings." Given the examiner-focused policies included in the EPQI and its intended goals, whether examiners actually have a significant influence on eventual outcomes and uses of patents is an important policy question.[11] We show empirically that patent examiners at the United States Patent Office indeed have a large causal impact on how their granted patents are used within the intellectual property system. Their impact is evident both through the "extensive margin" of accepting or rejecting patent applications and through the "intensive margin" of forcing modifications to patent applications during the examination process, which includes a "back-and-forth" between the applicant and the examiner. More broadly, our evidence shows the importance of the micro-determinants of intellectual property: important patent outcomes like litigation largely depend on examiner behavior and not simply on macro-determinants of the IP system such as the statutes in Title 35 of the US code.

Our research design starts from the fact that patent applications are conditionally randomly assigned to examiners.[12] Specifically, within an art unit[13], applications that are not continuation applications are randomly assigned to examiners. Leveraging this fact, we adapt the methodology used in the teacher value-added literature to estimate the causal impact examiners have on the probability that their granted patents are litigated or purchased by

---

[10]http://www.uspto.gov/blog/director/entry/enhanced_patent_quality_initiative_moving

[11]This paper is thus related to a growing literature in public economics that studies how to improve the management and retention of employees in the public sector, in particular in the context of teachers (e.g. Chetty et al., 2014).

[12]This fact has been discussed in Lemley and Sampat (2012). We have conducted our own interviews with patent examiners and the novely of our approach is to exclude continuation applications (including continuation, continuation-in-part, and divisional applications). This is an important adjustment for NPEs' portfolios, which have a high share of continuation patents, as documented in Section 3.3.

[13]Art units are small working groups, typically composed of about twenty examiners processing patent applications in similar technology classes.

NPEs. The intuition underpinning this research design is as follows: start from a patent outcome such as NPE purchase, calculate the share of an examiner's granted patents that feature this outcome, and compare this share across examiners working in the same year and art unit. If there is large (and sustained) heterogeneity in this share across examiners for a particular outcome, then our methodology will yield a large causal effect of each examiner for this outcome. We find that there is significant variation in the examiner causal effects for both litigation and NPE purchase outcomes. The signal standard deviation of the examiner causal effect distribution for patent assertion by NPEs corresponds to about 62% of the baseline patent assertion rate. The corresponding number for NPE purchase is 51%. In contrast, it is much lower for purchases by regular entities, around 14%. This provides initial evidence that examiners do have a significant influence on important IP market outcomes, and that the patent acquisition behavior of NPEs differs from that of regular firms.

As a second step, we look for mechanisms that explain the examiner variation in each of these outcome variables. To do so, we calculate the causal effects of each examiner for various behaviors observed during prosecution, such as their tendency to edit the text of the application and to cite various patentability criteria when rejecting a patent application. Using detailed prosecution data in this way allows us to measure whether a given patent examiner typically clarifies, narrows or rejects patent applications, which in turn allows us to infer the whether the patents granted by this examiner tend to be "weak." Then, we use leave-one-out versions of these examiner causal effect to predict the actual outcome of a given patent, such as NPE purchase and assertion in litigation. Another way to view this exercise is from a patent statistics perspective: we are projecting examiner characteristics onto the patents that they grant, and then comparing these randomly assigned characteristics across patents that are or are not purchased by NPEs. We discuss in the main text the selection effects inherent in this research design (some outcomes are observed only conditional on the patent application being granted).

We find that patents purchased by NPEs are, on average, granted by examiners who allow more incremental patents and patents with vaguer language. This result also holds for the subset of patents that are asserted by NPEs in litigation. As a comparison, practicing entity purchasing behavior exhibits very little dependence on the tendencies of examiners. An interesting point to note is that patents litigated by practicing entities also appear to be more incremental and vaguer relative to others in their technological cohort, but the effect sizes are much smaller than those exhibited by NPE litigated patents.

Taken together, our results suggest that NPEs purchase and assert different types of patents than regular firms, and appear to engage in some degree of rent seeking by purchasing patents for reasons orthogonal to or even negatively correlated with technological merit and social value. In ongoing work, we are looking for additional evidence that NPEs tend to buy and assert "weak" patents by studying whether the examiners that grant NPE patents tend to be reversed during appeal procedures at the patent office or invalidated in non-NPE lawsuits.

In terms of policy, our results suggest that examiner-focused reforms at the USPTO, such as examiner training or hiring more examiners, may have high social returns in terms of limiting the supply of inputs desired by NPEs. The lower-bound cost estimates we obtain in the paper, albeit in a partial equilibrium framework, suggest that these policies may represent excellent public investments, with expected returns close to 80%.

There are a few caveats and limitations to our study, but we believe they do not alter the general interpretation of our results. First, we only analyze observable outcomes in the IP system such as patent transfer and patent litigation, which do not encompass other important outcomes such as demand letters and licensing. Second, we may not have found all patents belonging to NPEs, given their tendencies to obfuscate their portfolios using shell companies, but this would only weaken our quantitative results. Finally, the nature of our methodology means that we are unable to make judgments on the behaviors of specific NPEs if they have small patent portfolios. This means that there may be NPEs that act more like IP intermediaries rather than rent seekers. However, given our findings, reforms to the patent examination process would actually serve as a sharp policy instrument, affecting only NPEs that rely on vague and obvious patents.[14]

The remainder of this paper is structured as follows. Section 2 provides background on the patent examination process, non-practicing entities, and patent and USPTO policies. Section 3 presents summary statistics and preliminary evidence based on the data we collect on NPE patent purchases and USPTO examiners. Section 4 describes the empirical framework and presents the baseline results on the size of the examiner causal effects. Section 5 conducts heterogeneity analysis to distinguish betwen various mechanisms. Section 6 discusses policy counterfactuals using our computed results. Section 7 concludes and discusses directions for future research.

3.2. Background

3.2.1. USPTO Examination Process

This paper relies extensively on the data from and the features of the patent examination system at the USPTO, which we describe briefly here.[15] When the USPTO receives a patent application, it first sends it to a vendor to classify the application based on technological area. Based on this initial classification, the application is then assigned to one of around 600 art units, which are comprised of groups of examiners with similar technological expertise. Once assigned to an art unit, the patent is assigned to examiners either based on examiners' current workload or based on the last digit of the application (Lemley and Sampat (2012)).

[14] As we will discuss later, such reforms would be beneficial even if NPEs served as efficient screening intermediaries that sift through patents with vaguely worded claims for ones with technological value. Reforms would deal with the vague claim problem at the source rather than having NPEs fund their screening through costly litigation.

[15] See Appendix B of the Sampat and Williams working paper for a more detailed explanation. See Lemley and Sampat (2012) for a detailed exploration of the assignment system.

As discussed in greater detail later in the paper, this assignment process is as good as random within art units.

Once the application has been assigned to an examiner, it is then evaluated by the examiner to ascertain whether it meets all the criteria for grant. If it does, then the examiner grants the patent immediately. According to USPTO figures, the percentage of patents granted right away ("first action") is around 10% only (Carley et al (2014)). Typically, the examiner issues either a final or a non-final rejection, also known as blocking actions, citing all of the eligibility criteria that are not met by the patent application. After rejections of either type, the applicant can abandon the application or send a response to the patent office, often editing the text of the application in order to address the criteria that the patent examiner deemed unmet. This then starts a back-and-forth process between the examiner and the applicant until the application is either approved or abandoned. In our analysis, we focus on applications with either of these outcomes, also known as disposed applications. The nature of the examination process described above shows that there is a possibility for examiners to differ not only in terms of the rate at which they grant patent application (as has been extensively discussed in the existing literature using "examiner instruments"), but also in terms of how they affect the nature of the eventually granted patent (breadth, clarity, etc.).

Examiners can affect the nature of the patents they grant through the aforementioned blocking actions. Based on the criteria cited in a rejection, an applicant will accordingly edit the text of the patent application before sending it back to the examiner. The main patentability criteria that examiners cite when blocking an application are the following sections from Chapter 35 of US Code:

1. Section 101 (Patentability and Utility): The patent needs to satisfy eligibility requirements and have the potential to be useful.

2. Section 102 (Novelty): The patent's claims are not covered in the prior art or in older

academic literature.

3. Section 103 (Non-obviousness): The difference between the invention and prior art is significant enough that it would not have been obvious to a "person having ordinary skill in the art."

4. Section 112, paragraph one (Sufficient disclosure/"Enablement"): The application "shall contain a written description of the invention... in such full, clear, concise, and exact terms as to enable any person skilled in the art ... to make and use [the invention]."

5. Section 112, paragraph two (Claim clarity/"Definiteness"): "The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention."

We show later that there is consistent and substantial variation in the usage of these provisions across examiners within the same art unit in any given year. The main focus of this paper will be on Section 103 and Section 112, paragraph two (also known as 112(b)), because these two blocking actions rely on the subjective judgement of the examiner, as reflected by the adjectives used to define them ("obvious" and "clear"), and because they can be used to test for the rent seeking behavior of NPEs.

One last element relevant for our discussion is the aforementioned claims of a particular patent. As noted in Section 112(b), claims in an application attempt to summarize and delimit the boundaries of a particular invention. The interpretation of text in a patent's claims is often the main focus of patent litigation. One concern with unclear claims text is that there will be many interpretations, and therefore a higher likelihood of confusion and disagreement over whether a product infringes on the patent. Examiners who are more likely to issue 112(b) blocking actions will on average grant patents with clearer claims text.

3.2.2. Relationship to Existing Literature

This paper builds on and contributes to several literatures. First, there is a growing economic and legal literature on NPEs. Cohen, Gurun, and Kominers (2014) investigate the characteristics of the defendants in NPE litigation. They find that NPEs tend to target firms with positive cash flow shocks, even if the cash flow shock hits a subdivision of a conglomerate that is different from the one accused of infringement. They also document other characteristics of firms targeted by NPEs such as number of lawyers, which provide evidence that NPE behavior is unrelated to actual infringement. Our paper complements their analysis by focusing on the characteristics of the key inputs in the NPE production function, namely patents. Despite their differences in focus and methodology, in line with our results they find that the behavior of NPEs is driven by factors unrelated to conventional use of the IP system, which can be thought of as purchasing patents based on technological merit and litigating based only on patent infringement.

Another strand of the NPE literature that we rely on is the classification of NPEs, which has generated widespread disagreement within academia and policy-making circles.[16] One simple approach is to apply an NPE label to any entity that makes all or most of its revenues from licensing and litigation. However, this broad definition would apply to technology development companies, some university-based IP entities, and failed start-ups, entities that many consider to be important to the innovation system. To address this issue, a recent paper by legal scholars Cotropia, Kesan, and Schwartz (2014) manually classifies all plaintiffs of IP lawsuits in the years 2010 and 2012, assigning each plaintiff to one of eight categories.[17] Their categories attempt to distinguish between possibly rent-seeking patent holding companies

[16] Including the recent debate in Congress over how to define the entities which the proposed fee-shifting provisions in the Innovation Act (H.R. 3309) would apply to.

[17] University/College, Individual/family trust, Large aggregator, Failed operating company/failed start-up, Patent holding company, Operating company, IP Holding company of operating company, and Technology development company.

and large patent aggregators on the one hand and the aforementioned "benevolent entity types" on the other hand. We make use of multiple sources in generating our NPE portfolios, and ascertain that our core results are robust to various portfolio construction methods.

In addition to Cotropia et al. (2014), there are several papers in the legal literature that provide descriptive evidence of NPE behavior. Fischer and Henkel (2012) show that patents acquired by NPEs are classified under more technology classes and received more citations, traditional measures of patent value. An earlier study, Allison, Lemley, and Walker (2009), considered patents that were litigated multiple times, and also found that they received more citations and were most likely to be owned by NPEs. These broad patterns also appear in the data we have collected, but our approach addresses possible endogeneity problems in citation statistics, such as the "publicity effect" of litigation on citations discussed by Lanjouw and Schankerman (2001). Our analysis suggests that in our specific setting, citations and independent claims may not be a valid indicator of social or technological value.

On the USPTO side, there is a literature focusing on the patent examination process. As mentioned earlier, we use the random assignment of applications for our identification strategy, a feature also exploited by Sampat and Williams (2015) in their investigation of follow-on innovation in human genome patents. Their and our work build on Lemley and Sampat (2012), which studies the issue of random assignment in detail by interviewing USPTO examiners. They found that random assignment within art units is plausible. Our contribution here is twofold. First, we show the importance of removing continuation applications, otherwise IV estimators are biased because continuation applications are not randomly assigned. Second, we show that examiners have an effect beyond the decision of granting the patent: they affect the nature of the granted patent (breadth, clarity, etc.), which is important to consider when gauging the validity of the exclusion restriction in any IV framework based on examiner assignment.

Beyond the issue of random assignment, the USPTO literature also discusses the validity of issued patents. There is a growing consensus, reflected in recent legislation, that US Patent Office examiners issue many patents of dubious validity. There is a debate as to whether this is primarily due to examiners' inability to distinguish these from other valid applications or to institutional constraints that make it very difficult for examiners to reject patents. Lei and Wright (2009) document that examiners distinguish weak patents from others that are stronger and, bearing the burden of proof of non-patentability, search more intensively for prior art that might bolster a case for rejecting weaker patents. They conclude that USPTO rules and procedures induce informed examiners to grant many of these weak patents. In contrast, we find that examiners have a large effect on the nature of the patents they grant and that their behavior does not appear to be significantly constrained by the existing USPTO rules and procedure.

Previous research has established that although patent examiners are charged with a uniform mandate, in practice examiners have a fair amount of discretion, and this discretion appears to translate into substantial variation in the decisions different examiners make on otherwise similar patent applications (Cockburn et al., 2003; Lichtman, 2004; Lemley and Sampat, 2010, 2012). Our research builds on the basic strategy used in Cockburn et al (2003) - we compute examiner effects with newer methodology to generate reliable magnitudes, analyze a different setting (NPEs and patent assertion), and add the extra step of correlating examiner effects with detailed data on the application process that was not available to earlier researchers.

Our methodological approach is adapted from the teacher value-added literature (Kane and Staiger (2008), Chetty et al. (2014)). Our paper is the first, to our knowledge, to apply this methdology in the setting of intellectual property. In addition, we discuss methodological issues in the case of rare and binary outcomes. Our methodology is also related to the examiner or judge leniency instrumental variables approach, used for instance by Sampat and Williams (2015) in the context of patent examination to analyze follow-on innovation. We discuss this link in greater detail in Appendix D. An additional methodological contribution here is to show that the validity of the instrument requires the exclusion of continuation applications.Our focus is on the "intensive margin", namely the effect an examiner has on a patent conditional on grant. But we also show results in the appendix using the examiner leniency methodology, which encompasses both intensive and extensive margins. The motivation for focusing on the intensive margin is that it allows us to move beyond just the effect of approval on outcomes and analyze changes in the nature of the patent.

Finally, our paper indirectly contributes to the vast patent statistics literature. Key papers in this area include Pakes (1986) and Hall, Jaffe, and Trajtenberg (2001), which look at patent value in the context of optimal renewal decisions and forward citations, respectively. Other papers in the literature investigate the role of scope (Lerner (1994)) and the number of independent claims as markers of patent value. Our methodology introduces a new set of randomly assigned patent statistics based on the projection of examiner tendencies onto the patents that they grant, which avoids endogeneity issues in some of the aforementioned widely-used statistics.

3.3. Data and Descriptive Statistics

3.3.1. Data Construction

We combine several data sources for our analysis. The first is patent data on both granted and ungranted applications. The American Inventors Protection Act of 1999 stipulated that patent applications filed on or after November 29, 2000 would be published eighteen months after the filing date. Prior to the legislation, the USPTO only published application information for granted patents. We make use of the Patent Examination Research Dataset compiled by the Chief Economist Office at the USPTO, and discussed in Graham, Marco, and Miller (2015). This covers applications in the period from November 2000 through December 2014. In our data, we see about 12,000 unique examiner names, and 650 unique art unit codes. We also make use of the patent data collected by Lee Fleming and co-authors [18], which contains organized extracts of the patent information made available by Google, and process the data using their name matching algorithm to assign inventor and assignee identifiers to each patent. Finally, we also process the aforementioned organized extracts from Google to obtain additional information on non-patent citations and IPC/CPC patent classifications.

In addition to basic patent application and grant information, we also make use of application-level blocking action data collected by Frakes and Wasserman (2015), which they have generously shared with us. As discussed earlier, data on blocking actions will give us added insight into the nature of patents granted by each examiner, and has, to our knowledge, not been used before Frakes and Wasserman (2015).[19] The Frakes and Wasserman data covers all applications filed after January 2001 and that are disposed by July 2012, which is only part of the sample available through Patent Examination Research Dataset. The applications in this set, which will serve as our core analysis sample, covers about 1.9 million disposed applications and 1.27 million granted patents (versus 2.68 million disposed applications and 1.82 million granted patents in the Patent Examination Research Dataset). We perform various robustness checks on the unrestricted sample.

Second, we collect data on NPE patent portfolios. The starting point of our exploration centered around the patent portfolio of Intellectual Ventures, a prominent NPE. Intellectual Ventures holds an estimated 25-30k US patents, and released a list of around 20k on their website in November of 2013. As we note later, some of these patents are applied for by Intellectual Ventures, and we exclude them in our formal analysis. To augment this data,

[18] The data is from the Patent Database Search Tool available at http://rosencrantz.berkeley.edu/

[19] This data is also collected by IP service websites such as Juristat, an indication of the important role of examiners in the patent application process.

we constructed NPE lists using two data sources on NPE names. Our main source is a set of entity names shared with us by RPX, a defensive patent aggregator that monitors NPE activity. The list of entities identified by RPX is similar to the one used in Cohen et al. (2015), because RPX recently purchased PatentFreedom, the source of their data. However, an important point to note is that we exclude individuals and universities from the list used by RPX in their reports.[20] A second source for NPE names comes from Cotropia et al (2014).[21] The limitation to this approach is that their paper only classifies plaintiffs of IP lawsuits in the years 2010 and 2012, so it would miss entities that are only active in other years. Both lists are then matched to assignee name in the Patent Assignment Dataset recently constructed by Alan Marco, Amanda Myers, and their colleagues at the USPTO and discussed in Marco et al. (2015).

Third, we use a combination of sources to look at lawsuits involving patents. This includes data from LexMachina, Darts IP, and RPX, three organizations that track intellectual property lawsuits and NPE activity. Our NPE lawsuit indicator is also derived from RPX classifications. This follows the approach taken by Cohen et al. (2015). Each of these sources tracks intellectual property lawsuits since 2000, which is ideal for our analysis because we start observing abandoned applications in November of 2000. In addition, LexMachina tracks patents that are eventually appealed through inter partes review, which was instituted as part of the America Invents Act.

Fourth, we collect additional characteristics on each published application and granted patent. This includes the text in the claims section of both applications and granted patents, backwards citations by who added it [22], assignment information by type of event, and maintenance fee payment information. For performing robustness checks, we also collect additional

[20]We additionally exclude Wisconsin Alumni Research Corporation and Children's Medical Center Corporation, which RPX classify as companies.
[21]Publicly available at npedata.com
[22]Either the examiner or the applicant. This data is available back to 2001.

patent-level indicators for special types of patents, including ones purchased by regular firms, assigned to universities, and listed in the FDA Orange Book.

3.3.2. Summary Statistics

In this subsection, we provide some summary statistics based on the data sources described above, in order to offer a clearer picture of the patent examination process at the USPTO and the types of patents litigated and purchased by NPEs. Throughout the section, we will exclude continuation applications from our summary stats, as our core analysis will do the same in order to maintain random assignment. We include summary statistics based on all applications in the Appendix.

3.3.2.1. Art Units and Examiners

We start by describing various patterns in our USPTO examination data. We present the basic structure of art units at the USPTO and document the amount of variation across art units and examiners for various variables, particularly NPE purchases and indicators for various types of examiner behavior.

Overall, there are 670 art units in our dataset. 559 of these art units have at least one granted patent ending up in an NPE portfolio (around 83%). Art units exhibit significant variation in attributes. There is a long right tail for most statistics, and with the level of NPE activity having the highest spread. The top art units in terms of NPE Patent Rate are in the areas of Data Processing (3621, 3622, 3685, 3688) and Communications (2631, 2637). Detailed summary statistics at the art unit level are shown in Appendix A. In this section, we will focus on statistics at the art unit by application filing year level, which is the level at which there is random assignment, and patent examiners, who create the variation used in our study.

First, we look at the statistics at the art unit by year level, which is the level at which applications are randomly assigned to examiners. There are NPE purchased patents in about 50% of the 7,156 art unit by year data points. The results are displayed in Table 3.1. We also look at blocking actions on patent applications within each art unit and year, including appeals to requirements listed in Section 101 (patentability and utility), Section 102(a) (novelty), Section 103(a) (non-obvious), and Section 112(b) (clarity in claims text). The use of 103(a) and 112(b) appear to be somewhat frequent, although 112(b) has a higher degree of variation across art units.

Table 3.1.: Art unit by year level statistics

Art unit by year level statistics	Median	Mean	Standard Deviation	Max
Examiners	15	15.5	15.0	155
Cases Processed	142	331	461	4510
Patents Granted	69	221	328	2538
NPE Patents	0	2.26	6.64	129
Grant Rate	0.64	0.62	0.20	1
NPE Patent Rate	0.001	0.011	0.020	0.333
Use of Section 101	0.029	0.097	0.132	0.667
Use of Section 102(a)	0.014	0.020	0.022	0.2
Use of Section 103(a)	0.47	0.45	0.20	0.96
Use of Section 112(b)	0.20	0.21	0.13	0.77

Notes: This table summarizes the statistics computed at the art unit by application year level, weighting each art unit by year equally. NPE patents are identified using the RPX list of entities, using the routine discussed in Section 3.3.1. Grant rate refers to the fraction of disposed applications (as of December 2014) that are granted in a given art unit by year. The NPE Patent rate refers to the fraction of granted patents in an art unit by year that have been purchased by NPEs (as of December 2014). "Use of Section" variables refer to the sections in Chapter 35 of US Code (discussed in Section 3.2.1) that are cited in examiner blocking actions.

Next, we show the statistics associated with patent examiners in Table 3.2. This gives us a basic idea of the amount of variation across examiners. There are 12,032 unique examiners, and 98.7% of the examiners are at one point assigned to an art unit with some NPE activity. There is significant variation amongst examiners in all of the computed statistics, with the

NPE rate having the highest spread relative to the mean. A major part of this variation comes from across art unit differences, as discussed above. In addition, there is also noise in the sets of applications that examiners receive. In the remained of the paper, we use various approaches to eliminate the noise and identify the signal variance coming from examiners: we find that it is still substantial.

Table 3.2.: Examiner level summary statistics

Examiner Level Statistics	Median	Mean	Standard Deviation	Max
Cases Processed	119	190	215	1600
Art Units	2	1.80	0.96	7
Years in Data	7	6.35	3.19	12
Patent Grant Rate	0.59	0.57	0.23	1
NPE Patent Rate	0	0.011	0.023	0.5
Use of Section 101	0.029	0.092	0.129	0.738
Use of Section 102(a)	0.010	0.019	0.025	0.240
Use of Section 103(a)	0.49	0.47	0.17	0.95
Use of Section 112(b)	0.19	0.21	0.13	0.76

Notes: This table summarizes the statistics at the examiner level, weighting each examiner equally. Rate computations are restricted to examiners with more than 20 cases in the data. The variables here are computed in the same way as the ones shown in Table 3.1.

3.3.2.2. NPE Purchased Patents

In this part, we explore these characteristics of patents purchased by NPEs, in order to understand NPE purchasing tendencies and the technological areas and industries in which they are the most active. Once again, we exclude continuation applications from our main analysis, and report overall numbers in the Appendix. The main findings here are that NPE-purchased patents are predominantly concentrated in computer hardware and software, and NPE patents have very different pre- and post- examination features relative to the average patent.

We start by looking at the patent classes in which NPEs are the most active. In Table 3.3, we display aggregate counts of all patents held by NPEs, at the NBER category and sub-category level (Hall et al (2001)). This confirms the general view that many of the patents are concentrated in IT, although there are a small number of patents in other fields. As a comparison, we have also computed similar statistics for non-NPE asserted patents in Table C.3, and found a very different composition of categories. Of course, this evidence alone does not help differentiate between the two theories on NPE behavior, as it could be that the most resource-constrained firms are in IT.

Table 3.3.: NPE Patent Holdings by NBER Technology Category

Panel A: Primary Technology Categories

NBER ID	Category Name	Patents
2	Computers & Communications	7,587
4	Electrical & Electronic	3,050
5	Mechanical	684
-	*New Classes (since 2001)*	629
6	Chemical	415
1	Others	392
3	Drugs & Medical	198

Panel B: Secondary Technology Categories

NBER ID	Subcategory Name	Patents
21	Communications	3,368
22	Computer Hardware & Software	2,868
46	Semiconductor Devices	1,147
24	Information Storage	900
-	*New Classes (since 2001)*	629
41	Electrical Devices	496
23	Computer Peripherials	451
49	Miscellaneous-Elec	411
45	Power Systems	359
42	Electrical Lighting	352
54	Optics	294
19	Miscellaneous-chemical	282
69	Miscellaneous-Others	231
59	Miscellaneous-Mechanical	211
43	Measuring & Testing	159

Notes: the NPE patents are identified based on RPX classifications (see Section 3.3.1). We take the primary USPTO technology class for each patent, and then aggregate up to subcategory and category levels, using the NBER patent crosswalk provided by Hall et al. (2001).

Next, we look at the sources of patents owned by NPEs. To do this, we look up the original assignee on each NPE patent in our list. The results of this exercise are shown in Table 3.4. The plurality of the roughly 35k NPE owned non-continuation patents are initially unassigned. In addition, some patents are initially assigned to entities associated with Intellectual Ventures, such as The Invention Science Fund and Searete. Since the focus of our paper is on patent acquisition, we will not count these patents as NPE-purchased for our formal analysis.[23] Another interesting point to note is that while some of these patents were originally assigned to companies that have gone through bankruptcy (Eastman Kodak and Polaroid Corporation), others were initially owned by firms that have not, such as GE, Lucent, and Micron. As we discover later in our analysis with initial-assignee fixed effects, firms with many granted patents also appear to selectively sell their vaguer and more obvious patents to NPEs, so our effect is only partly driven by cross-firm differences and individual inventors.

Table 3.4.: Aggregate NPE portfolio patent counts by initial assignee

Initial Assignee	# Patents	Initial Assignee	# Patents
(Unassigned)	1,434	Harris Corporation	104
Eastman Kodak Company	806	MIPS Technologies, Inc.	93
Micron Technology, Inc.	381	NEC Corporation	84
Telefonaktiebolaget L M Ericsson	334	MOSAID Technologies	83
Nokia Corporation	322	DaimlerChrysler AG	79
Koninklijke Philips Electronics	302	NEC LCD Technologies	76
Matsushita Electric Industrial Co, Ltd.	256	Empire Technology Development	71
NXP B.V.	230	Lucent Technologies Inc.	70
Panasonic Corporation	195	Raytheon Company	69
American Express Travel Related	167	University of California	68
Global OLED Technology LLC	146	DPHI Acquisitions, Inc.	67
Hynix Semiconductor Inc.	127	Lite-On Technology Corp.	66
Industrial Technology Research Institute	125	Cypress Semiconductor Corporation	65
Virginia Tech Intellectual Properties	120	*Searete LLC*	64
The Invention Science Fund I, LLC	111	LG Electronics Inc.	60

Notes: We use assignee names from the Lee Fleming database. Italics indicate entities associated with Intellectual Ventures (patents granted directly to them). For entities other than Intellectual Ventures, we include only patents that are assigned to them, but not as part of an "employer assignment," as classified in the USPTO assignment dataset.

Finally, we compare basic statistics of NPE-purchased patents with other patents in their

[23] As detailed in the appendix, the rest of the portfolio is constructed excluding patents that are initially assigned to NPEs on our lists. On a different but related note, we verify that Intellectual Ventures patent applications appear to be randomly assigned to patent examiners, which is consistent with our identification assumption.

technological cohort. The main takeaway from this exercise is that patents purchased by NPEs have different initial and post-examination characteristics relative to other patents in the same cohort.

To construct a crude control group, we take all granted patents that had the same application year and art unit assignment as an NPE patent, and re-weight the control patents based on the number of NPE patents in their cohort. The statistics we compute include ex-ante characteristics, such as the number of independent claims in the initial application, examination characteristics, such as the frequency of blocking actions and changes in words in the primary claim, and ex-post outcomes. such as forward citations and re-assignment events.The results are shown in Table 3.5. The statistics presented here are robust to the NPE list we use.[24]

[24]Including removing all Intellectual Ventures patents to look at the portfolio of the rest of the NPEs and looking at the Kesan list of patent holding companies and large aggregators.

Table 3.5.: Summary Statistics: NPE vs. Non-NPE Patents

Statistic	NPE Patents	Control Patents	Difference
Initial Application Characteristics			
Independent claims at Application	5.00	4.07	0.92***
Examination Outcomes			
Examiner leave-one-out allowance rate	0.766	0.759	0.006***
Number of Examiner Rejections	1.93	1.87	0.06***
Change in # Independent Claims	-1.58	-0.83	-0.75***
Use of Section 101 - Lack of utility or eligibility	0.085	0.090	-0.005**
Use of Section 102 - Prior art exists	0.021	0.019	0.002
Use of Section 103 - Obvious invention	0.48	0.47	0.01
Use of Section 112(a) - Improper Disclosure	0.059	0.051	0.008***
Use of Section 112(b) - Vague Claims	0.18	0.17	0.012***
Post-grant Outcomes			
Citations	8.86	6.71279	2.14***
Number of re-assignments	1.67	0.36	1.31***
Date of latest re-assignment relative to grant (days)	1245	881	364***
Involved in IP Litigation	0.035	0.005	0.030***
Inter Partes Review Instituted	0.0042	0.0002	0.0040***

Notes: For each NPE patents, identified using the RPX list as discussed in Section 3.3.1, we take the set of non-NPE patents in the same art unit and application year cohort, and re-weight them so that there is the equivalent of one control for each NPE patent. Then, we compute average characteristics and outcome values for each group and also compute the difference in means. Examiner leave-one-out allowance rate for each patent is computed based on Equation 3.1. ** p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01*

The results show that there are significant differences between NPE-purchased patents and others in the same year and technological cohort. In terms of characteristics at the start of application, the set of NPE-purchased patents have a higher number of independent claims at application. This suggests that the applications aim to claim a broad scope, but as we see in the average number of changes in independent claims, many of these are removed before final grant.

Next, the examination indicators provide a mixed message. NPE-purchased patents come from examiners with a slightly higher leave-one-out grant rate, although this is a noisy measure of true examiner approval rates. In addition, these patents tend to have more examiner rejections ("final" and "non-final") and more independent claims removed. There is also significantly more 112(b) usage on the NPE-purchased patents, although this could suggest either a vaguer application to begin with or examination by more stringent examiners.

Finally, in terms of eventual outcomes, NPE-purchased patents are more likely to be cited. Citations have been the primary measure of patent value in the literature, but as pointed out earlier, part of the difference here could be from an exposure effect.[25] NPE-purchased patents are also more frequently re-assigned, although this is partly mechanical, as we construct this group based on re-assignment data. Furthermore, conditional on re-assignment, NPE patents are more likely to be purchased later in a patent term than other purchased patents, a result similar in nature to Love (2013), which finds that NPEs assert patents late in their term. These patents are also more than seven times more likely to end up in litigation, and twenty times more likely to be instituted under the new Inter Partes Review system.[26] These post-grant outcomes are where the two groups of patents differ most, but some of this is unsurprising given that the groups are defined based on a post-grant outcome, namely purchase by an NPE.

In general, the simple comparisons are hard to interpret, as they reflect both the nature of the original application and the changes made by the patent examiner, which sometimes create offsetting differences. Of course, we could look to improve the matching criteria to create a better control group, but this is not always feasible because of the limited number of patents within each art unit cross year cohort and hard-to-measure quantities such as vagueness of the initial claims language. Instead, as we explain in our methodology, we

[25] One particular way this could work is through the threat of "inequitable conduct" rulings. A patent holder can have his patent ruled unenforceable if he can be shown to have deliberately not cited prior art in his patent application. An applicant can avoid this risk by citing all widely known patents that are involved in lawsuits or are in the news.

[26] Instituted cases represent appeals where the petitioner has a "reasonable likelihood" of prevailing. We use this outcome as there have been very few cases with final verdicts under this new system put in place after the America Invents Act.

indirectly control for differences across applications by looking for consistency of examiner differences over time, leveraging random allocation.

3.3.3. Examiner Variation

Next, we present evidence on random assignment of patent application to examiners and we show preliminary evidence on the heterogeneous effects examiners have on patent outcomes.

To show preliminary evidence on the heterogeneous effects examiners have on patent outcomes, we rank examiners by their allowance rate and blocking action usage, and look for variation in the fraction of litigated and NPE purchased patents across the quartiles. Specifically, for each granted patent, we compute an examiner leave-one-out allowance rate, excluding that particular case from the calculation.

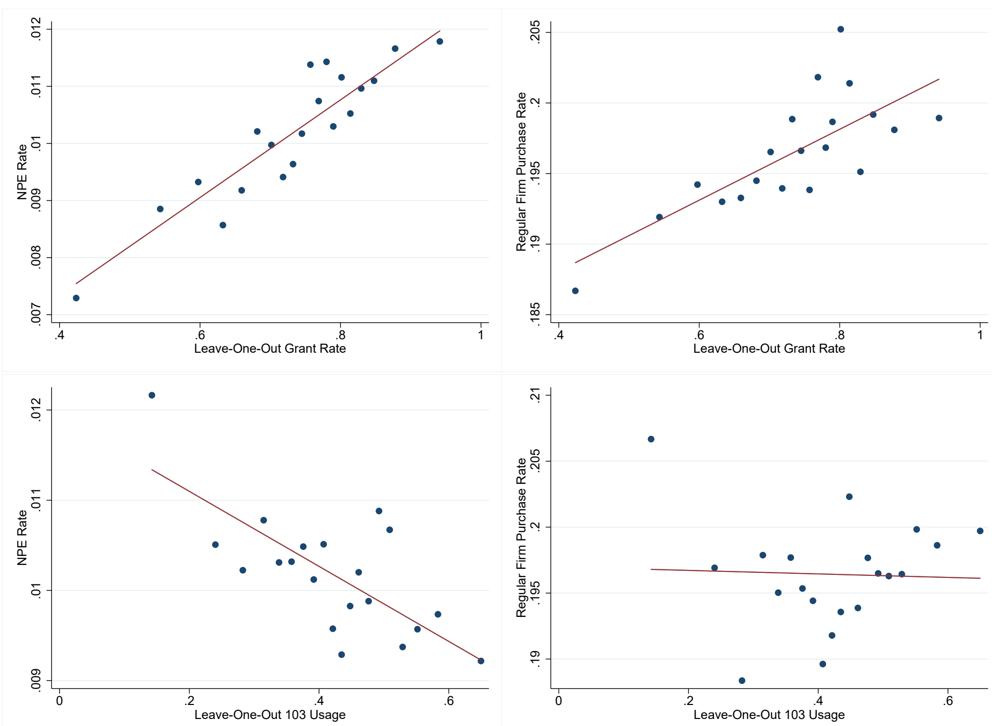
$$Leave - one - out\ Allowance\ Rate_i = \frac{\sum_{k \neq i, k \in J(i)} GrantDecision_k}{N_{J(i)} - 1} \tag{3.1}$$

where i is the granted patent of interest, $J(i)$ is the set of (granted and ungranted) patent applications processed by the examiner who granted patent i, and $N_{J(i)}$ is the number of applications in $J(i)$. A high value for this variable indicates that patent i was granted by an examiner who grants a high share of the patent applications they process. We use a similar approach to compute leave-one-out blocking action usage.

Figure 3.1 shows the relationship between examiner leave-one-out means over their whole career and subsequent granted patent outcomes, such as NPE purchase and non-NPE purchase. The NPE purchase outcome shows a strong positive association with examiner leave-one-out grant rate, a relationship that is much weaker for the non-NPE purchase outcome. In addition, NPE purchase is negatively associated with examiner usage of 103(a) blocking actions, whereas non-NPE purchase is uncorrelated.

One main problem with pooling data from an examiner's entire career is that we could be capturing secular technological variation over time. To check for this, we have also compute leave-one-out means at the art unit by year level, which is the approach taken by Sampat and Williams (2014). The results are reported in Figure C.1, and provide further evidence that patents granted by examiners with a higher allowance rate are more likely to be purchased by NPEs. In fact, the weak positive relationship between non-NPE purchase and grant rate disappears under this specification. Of course, neither piece of preliminary evidence addresses noise in the grant rate calculations and the consistency of examiner rankings across years, which we will address through our formal methodology in Section 3.4.

Figure 3.1.: Preliminary Evidence on the Relationship Between Examiner Tendencies and Patent Outcomes



Notes: Results are computed on the baseline sample (non-continuation granted patents covered by Frakes and Wasserman). Examiner effects are computed as a leave-one-out mean on their entire history. Outcome data is at the granted patent level (fraction conditional on grant). Results are computed absorbing art unit fixed effects, and restricting to examiners who review at least 50 applications.

In order to test the validity of the random assignment of examiner, we conduct placebo tests by regressing a number of outcomes that were determined before the time of patent application on the leniency of the examiner in charge of this patent application. The outcomes we consider are either defined at the inventor level or the patent level. Specifically, we consider three outcomes that are related to the "value" of a patent: the number of patents granted to an inventor in previous years (as a proxy for inventor quality), the number of independent claims at the time of publication of the application (a widely-used proxy for the quality of a patent) and the number of words in the first independent claim (another commonly used proxy for patent quality). In addition, we consider two outcomes that speak directly to the exclusion restriction: the number of patents of the inventor that were purchased by or issue to NPEs in previous years, and the number of patents of the inventor that were litgiated in previous years. We find no statistically significant relationship between these outcomes and the leave-one-out examiner allowance rate in our preferred sample, which excludes continuation application and repeated inventor-examiner pairs and adjusts the leave-one-out allowance rate for docketing time patterns.[27] However, if we keep the full sample of applications (which to the best of our knowledge is the current practice in the literature using patent examiner instruments), the placebo tests fails. Restricting the sample as we do is therefore key to preserve random assignment. These various results are reported in Appendix B.

Another key element of the research design is that our specifications are carried out conditional on the patent being granted: if an application is not granted, by definition we could not observe NPE purchase, litigation, etc... Therefore, the variation in the rate of NPE patents observed across examiners could be due to variation in the underlying quality of the pool of granted patents across examiners. If NPEs tend to purchase patents from a set of examiners with a pool of granted patents of high quality, then the interpretation that

[27]Continuation applications are not randomly assigned and should therefore be excluded. The rational for also excluding repeated inventor-examiner pairs (within the same artunit-year-class) is that some continuation applications may not be properly recorded. Applications that were submitted at the same time are likely to be assigned to the same examiner in a batch, therefore we compute the leave-one-out allowance rate only based on applications that were docketed in different months.

NPEs target "weak patents" would be misplaced. The underlying quality of a patent is not observed, but we find that the examiners who have a high rate of NPE purchases or litigation also have high allowance rates (as documented above) and do no narrow or clarify the claims of the patents as much as other examiners (as documented in Section3.5). In other words, average patent quality should be lower in the pool of patent granted by examiners with a high NPE rate. We test this hypothesis in Appendix B and indeed find support for it: inventors who get granted patents by examiner with high allowance rates tend to be less accomplished (i.e. they have been granted fewer patents in previous years). Therefore, our results across examiners are not driven by a selection effect on quality. Appendix B provides a more in-depth discussion of selection effects in our setting.

In the remainder of the paper, we first use a shrinkage methodology to estimate the causal effect of examiners on the probability that a granted patent becomes part of a NPE portfolio. We find very large effects, suggesting that the rate of patents in NPE portfolios could be reduced by over 50% by re-designing the patent examination process. Second, we systematically investigate the characteristics of examiners with a high NPE effect, studying all types of rejection during prosecution, changes to the claims and propensity to cite prior art. The results confirm the suggestive findings presented in this section: a large share of the patents that are part of NPEs' portfolio are "weak" patents, which were granted by lenient examiners and appear to be vaguely worded. The formal shrinkage and regression frameworks we use allow us to establish the statistical robustness of the preliminary findings here.

3.4. Estimating Examiner Causal Effects

In this section, we describe our overall research design, discuss threats to identification, and then show formal methodology and results on computing examiner causal effects.

3.4.1. Research Design Overview

Our research design starts from the fact that patent application are conditionally randomly assigned to examiners. Indeed, within an art unit, applications that are not continuation applications are randomly assigned to examiners.[28] The intuition of our research design is to look at the share of NPE patent (patents that are eventually purchased or asserted by NPEs) in the portfolio of granted patents of an examiner, compared with the share of NPE patents of other examiners working in the same year and art unit. If there is substantial heterogeneity across examiners, then we can conclude that the patent examination process plays an important role in the activities of NPEs and we can look at the characteristics of examiners who have a high share of NPE patents to learn about the important features of the patent examination process and infer the patent acquisition strategy of NPEs.

The first step of our analysis is to test for "excess variance" from examiners, i.e. do NPE shares differ significantly and consistently across examiners in the same art unit? To do so, we employ a shrinkage methodology filtering out the noise in the data. Intuitively, if an examiner granted only two or three patents, his estimated share of NPE patents will be extremely noisy. We address this in various ways, following the teacher value-added methodology in Kane and Staiger (2008) and then using a novel non-parametric Empirical Bayes methodology we introduce as a robustness check. This analysis is reported in the remainder of this section.

The second step, the focus of Section , is to estimate the correlation between the NPE examiner effect estimated in the first step and other examiner effects we estimate following a similar methodology (examiner allowance effect, examiner propensity to cite prior art, examiner propensity to use certain kinds of blocking actions, etc.) in order to learn about

[28]This fact has been discussed in Lemley and Sampat (2012). We have conducted our own interviews with patent examiners and the novelty of our approach is to exclude continuation applications. This is a very important adjustment for NPEs' portfolio, which have a high share of continuation patents, as documented in Section 3.3.

the mechanism at play.

The analogy between our setting and the teacher valued-added framework of Kane and Staiger (2008) is as follows: we treat examiners as teachers, applications within a given year by art unit as the student cohort, and we measure patent-level outcomes both at the time of grant, such as claims attributes, and eventual outcomes, such as ending up in an NPE portfolio and being asserted in IP litigation. This is analogous to the approach taken in various papers in the teacher effects literature, such as Chetty, Friedman, and Rockoff (2012), which uses test scores as the short-term indicator of teacher quality and wages at age 28 as the long-term outcome of interest. In their paper, the nature of the test scores is uncertain and higher wages are indicative of better outcomes. Here, we take the reverse approach: we use the short-term indicators such as 103(a) and 112(b) usage to better understand the nature of the long-term outcome, namely NPE acquisition.

3.4.2. Computing Examiner Causal Effects

The first step in our methodology is to compute an examiner causal effect. For our core results, we adapt methodology from the teacher value-added literature. This framework allows us to estimate an unbiased predictor of the effect that a given examiner has on a particular outcome for a given patent. This is done by taking the examiner fixed effect in a given year and shrinking it by a factor equal to the correlation of effects across years divided by the level of idiosyncratic noise in the outcomes, in order to extract the signal component in the data. These compute effects can then be treated as causal examiner effects, because of the random assignment structure detailed above.

More precisely, our empirical framework is as follows:

$T_{ijt} = X_{ijt}\beta + a_{ut} + v_{ijt}$

$v_{ijt} = \mu_j + \theta_{jt} + \epsilon_{ijt}$

where i indexes the patent, j the examiner, u the art unit and t the year. X_{ijt} represents controls for patent application characteristics, a_{ut} represents art unit by year fixed effects, to control for differences at the level of random assignment. μ_j is the examiner causal effect of interest (assumed to be constant over time), θ_{jt} represents an idiosyncratic examiner by cohort effect,[29] and ϵ_{ijt} is an idiosyncratic patent effect. We run Equation **??** using OLS to obtain the residuals. We then take a series of steps to compute examiner effect estimates using these residuals.

First, we compute the within examiner-year variance in v_{ijt} to obtain an estimate of the variance of the idiosyncratic component:

$\hat{\sigma}_{\epsilon}^2 = Var(v_{ijt} - \bar{v}_{jt})$

where$\bar{v}_{jt} = \frac{1}{n_{jt}} \sum_{i=1}^{n_{jt}} v_{ijt}$, an average of residuals within a year. This step serves as an important for later adjustments for background noise. In our context, this idiosyncratic component has significant variance, as there are many features of patent applications that we are unable to observe or measure.

Second, we use the covariance between the average residual in an examiner's portfolio in year t and year $t+1$ as an estimate of the variance in the examiner component:

$\hat{\sigma}_{\mu}^2 = cov(\bar{v}_{jt}, \bar{v}_{j(t+1)})$

where the covariance calculation is weighted by the number of patents granted by each examiner (n_{jt}). This variance component - which we refer to as the "signal variance" of the examiners - is a measure of the variation in examiner effects across examiners. Intuitively, if all the true examiner μ_j's are close to zero relative to the size of the outcome, we may still get variation in the calculated annual average residuals $\bar{v}_{jt}$ from the idiosyncratic error draws, but our methodology will then pick up a very low signal covariance across years, because the

[29]This could result from the examiner being more familiar with certain technology cohorts or just random fluctuations in examiner behavior over time. The key to using this framework is that θshould not be serially correlated, which would alter the interpretation of the calculated signal variance computed in Equation **??**.

idiosyncratic components are uncorrelated.

Third, the variance of the examiner-cohort idiosyncratic component is estimated as the remainder:

$\hat{\sigma}_\theta^2 = Var(v_{ijt}) - \hat{\sigma}_\mu^2 - \hat{\sigma}_\epsilon^2$

Next, we form a weighted average of the average residuals for each examiner in each year ($\bar{v}_{jt}$) that is a minimum variance *unbiased* estimate of μ_j for each examiner. Data from each year is weighted by its precision, with years in which the examiner granted more patents received more weight:

$\bar{v}_j = \sum_t w_{jt}\bar{v}_{jt}$

where

$w_{jt} = \frac{h_{jt}}{\sum_t h_{jt}}$

$h_{jt} = \frac{1}{\hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\epsilon^2}{n_{jt}}}$

The last step is to construct the "empirical Bayes estimate" of each examiner's effect by multiplying the weighted average of examiner-year residuals by a shrinkage factor:

$ExaminerEffect_j = \bar{v}_j \frac{\hat{\sigma}_\mu^2}{Var(\bar{v}_j)}$

where $Var(\bar{v}_j) = \hat{\sigma}_\mu^2 + (\sum_t h_{jt})^{-1}$.

The shrinkage factor is the ratio of signal variance to total variance, and is different across examiners based on cases examined. This final quantity, the examiner effect, has two desirable properties (see Appendix C for formal demonstrations of these properties). First, it has an empirical Bayes interpretation as the Bayesian estimate of the examiner effect, with a normal prior distribution centered around zero with variance equal to the signal variance calculated in the second step above.

Second, it also has a frequentist interpretation: the shrinkage factor represents the coefficient of a hypothetical regression of μ_j on $\bar{v}_j$. The regression coefficient is a ratio of the covariance of the two expressions, in this case $\hat{\sigma}_\mu^2$ because the other parts of $\bar{v}_j$ are uncorrelated noise, divided by the variance of the independent variable. Therefore, the estimated $ExaminerEffect_j$, although biased, has the same scale as the true examiner effect μ_j in this model, causing it to have a lower mean-squared forecast error relative to $\bar{v}_j$. We believe that the distribution of $ExaminerEffect_j$ better captures the scale of true examiner effect distribution relative to using $\bar{v}_j$, and therefore report results using the former in our main findings.[30] We later verify using a split sample approach that our examiner effect has this desired property.

One shortcoming of using this approach in our setting is that, as explained in Kane and Staiger (2008), a Bayesian interpretation of the results requires the assumption that the residuals of the regression are normally distributed, which may not be appropriate in our setting given that the outcome variable is binary. Nonetheless, we include controls, in particular art unit and year fixed effects, therefore the normality assumption is not necessarily violated. In addition, to address this this concern we also estimate the distribution of examiner effects using a binomial model, in the spirit of Ellison and Swanson (2010). The results are consistent across models, as detailed in Section 3.4 below.

3.4.3. Results

Here, we report the main results, computed using the Kane and Staiger framework. We report the distribution of causal examiner effects for our main outcomes of interest, and also report results on the signal variance parameter estimated as part of the methodology.

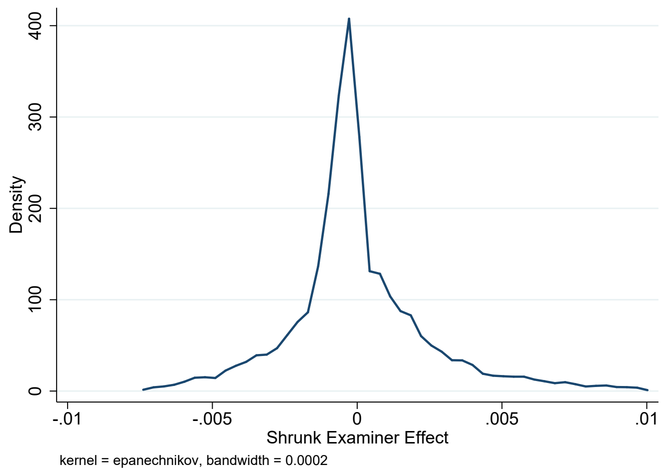
First, we plot the distribution of examiner effects on NPE purchase in Figure 3.2, keeping all examiners and weighting by the number of cases processed.[31] The baseline rate of NPE

[30] Using the distribution of $\bar{v}_j$ would leave to a higher variance.
[31] The results without weighting are very similar.

purchase is around 0.01, so the spread in the distribution accounts for a sizable fraction of this baseline rate.

To be more specific, we display in Table 3.6 provides various measures of the magnitude of these effects, expressed as a percentage of the share of NPE patents in the patent population. For each of these results, we also compute 95% confidence intervals using a bootstrap procedure, and report it in brackets in the line below the computed number. The first row reports the signal standard deviation, estimated as explained in section 3.4.2, which amounts to a staggering 52.95% of the NPE share. Under the Bayesian interpretation, this is the empirical Bayes estimate of the standard deviation of the distribution of true examiner effects $\hat{\sigma}_\mu$. Next, the standard deviation of the computed shrunk examiner effects distribution is around 24.33% of the baseline rate. This number is distinct from the signal standard deviation, as it reflects the distribution of posterior means from updating the prior for each examiner. The interquartile range of the shrunk examiner effect distribution is less than the standard deviation, still a sizable 7.25% of the NPE share. This suggests that the shrunk examiner effects distribution has excess density in the tails relative to a normal distribution. Finally, if all examiners above the 75th percentile of the distribution (i.e. examiners with an unusually high share of NPE patents in their granted patents) were replaced with examiners located exactly at the 75th percentile of the distribution, then the share of granted patents would decrease by 6.15%. Overall, these results indicate that patent examiners have a large causal effect on the probability that a granted patent becomes part of an NPE portfolio.

Figure 3.2.: NPE Examiner Effect Distribution



Notes: We run a kernel density plot of the distribution of shrunk examiner effects across all examiners, weighted by the number of granted patents for each examiner. The shrunk examiner effects are calculated using the methodology described in Section 3.4.2.

Table 3.6.: NPE Examiner Effect Distribution Characteristics

Measure of Examiner NPE Effect	Percentage of Sample NPE share
Signal standard deviation ($\hat{\sigma}_\mu$)	50.97% [33.7%,60.7%]
Standard deviation of Shrunk Examiner Effects	24.02% [12.26%,30.90%]
Difference between examiners at p25 and p75	6.86% [3.25%,8.51%]
Difference in distribution mean replacing examiners above p75 with examiners at p75	6.00% [3.02%,7.30%]

Notes: Bootstrapped 95% confidence intervals are displayed below the corresponding parameter estimate. The size of the parameters are normalized by the baseline rate of NPE-purchased patents in the sample (0.9%). The "difference in distribution mean" value is calculated by replacing shrunk examiner effects above the 75th percentile with the value at the 75th percentile, and then re-computing the average of the new distribution.

As a comparison, we also compute similar results for patents asserted by companies not on our NPE list. The baseline rate of these patents is 13.81%, and we display the results in Table 3.7. The major difference here is that the signal standard deviation is much lower, suggesting less heterogeneity across examiners in whether their granted patents end up being transferred on the IP market.

Table 3.7.: Non-NPE Purchased Examiner Effects Distribution Characteristics

Measure of Examiner non-NPE Effect	Percentage of Sample NPE share
Signal standard deviation ($\hat{\sigma}_\mu$)	12.01% [10.70%,14.47%]
Standard deviation of Shrunk Examiner Effects	6.32% [5.05%,7.32%]
Difference between examiners at p25 and p75	5.67% [4.47%,6.97%]
Difference in distribution mean replacing examiners above p75 with examiners at p75	1.15% [0.95%,1.35%]

Notes: Comparable to Table 3.6. Bootstrapped 95% confidence intervals are displayed below the corresponding parameter estimate. The size of the parameters are normalized by the baseline rate of non-NPE purchased patents in the sample (13.81%).

We also apply the same methodology on other patent outcomes and examination characteristics, such as the patent being purchased and litigated by NPEs, the patent being litigated by a practicing entity, or the use of various blocking actions during prosecution, which we will make use of in the discussion on mechanisms. Examiner effect distribution graphs for some variables are included in Appendix D. In particular, we find that examiner effects are quite large for 103(a) and 112(b) blocking actions (24.0% and 46.5%, respectively), which are known to rely more on the subjective judgment of the examiner. They are also large for the practicing entity litigation outcome (signal standard deviation of 62.1%), which we will explore further in the mechanism section.

3.4.4. Robustness Checks

3.4.4.1. Robustness of Basic Methodology

In this part, we report a series of robustness checks of the basic results presented in Section 3.4.3. Half are related to the analysis sample, and half are related to NPE classifications and specification. We find that the results are quite similar across these different settings. The results for the sample that includes continuation applications exhibit larger spread in NPE effects, as continuations are often assigned to the same examiner, and patents in the same family are likely to be purchased together. The Cortropia results may be a little weaker because of its lack of coverage, so we would be including actual NPEs that litigated in years other than 2010 and 2012 in the control group.

Table 3.8.: Robustness checks for the NPE examiner effect distribution.

Specification	Signal SD	Shrunk Effect SD
Baseline	62.8%	25.3%
Expanded Sample (PERD)	52.9%	24.3%
Include Continuations (PERD)	77.2%	48.1%
Expanded Sample (1976-2015)	70.6%	24.8%
Excluding Intellectual Ventures Patents	71.2%	33.5%
Cortropia et al. NPE Patents	43.1%	15.5%
Additional Patent Controls	54.9%	24.8%
Assignee Fixed Effects	33.9%	13.2%

Notes: Expanded Sample (PERD) refers to the sample of all non-continuation applications in the Patent Examination Research Dataset, which covers applications filed between 2001 and 2014. Expanded Sample (1976-2015) refers to the sample of all non-continuation granted patents in that time period in the data available from Google. Excluding Intellectual Ventures Patents refers to excluding them from the analysis, and looking at patents owned by other NPEs. Additional patent controls refers to adding the number of independent claims at application and first inventor experience as controls in the basic estimating equation (Equation **??**). All effects are normalized to the baseline NPE rate, which is different for the various specifications.

3.4.4.2. Empirical Bayes Count Model

To address the concern that the assumptions behind the Kane-Staiger (2008) methodology may not be satisfied, we use an alternative methodology. Specifically, we rely on an Empirical Bayes Beta-Binomial model, in the spirit of Ellison and Swanson (2010). This methodology allows us to set a more flexible prior on the distribution of examiner effects, and also allows us to more directly model the data generating process.

First, we aggregate data for each examiner j in year t and art unit a into the form (n_{jat}, r_{jat}), where n denotes the total number of granted patents for a given examiner and r the total number of NPE patents (or some other outcome) for this examiner. We then model the data generating process directly with a binomial likelihood on each data point: each examiner has some true probability p of having an NPE purchase a granted patent.

Next, we set up an empirical Bayes estimation of the prior of p. For each art unit in each

year, we start with a flexible prior distribution:
$p \sim Beta(\alpha, \beta)$

We can then form an integrated likelihood in order to estimate the hyperparameters α and β of the prior distribution:

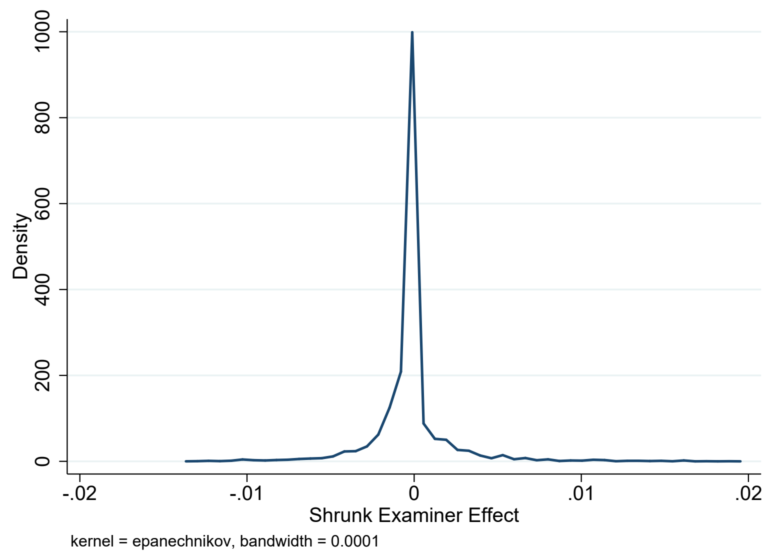
$$\begin{aligned}
L(r|n,\alpha,\beta) &= \int_{p=0}^{1} \begin{pmatrix} n \\ r \end{pmatrix} p^r(1-p)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp \\
&= \begin{pmatrix} n \\ r \end{pmatrix} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{p=0}^{1} p^r(1-p)p^{\alpha-1}(1-p)^{\beta-1}dp \\
&= \begin{pmatrix} n \\ r \end{pmatrix} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(r+\alpha)\Gamma(n-r+\beta)}{\Gamma(n+\alpha+\beta)}
\end{aligned}$$

where the second step just conjugates the inside to integrate to one based on the probability density function of the Beta distribution. We then pool the data points of examiners in the same year and art unit, and estimate the hyperparameters via maximum likelihood. These parameters allow us to calculate the equivalent of the signal standard deviation in the basic framework.

To compute the posterior mean for each examiner, we take the estimated parameters $(\hat{\alpha}, \hat{\beta})$ and updated them using the data: the posterior distribution is given by $Beta(\hat{\alpha} + r, \hat{\beta} + n - r)$. The mean of this distribution, $\frac{\hat{\alpha}+r}{\hat{\alpha}+\hat{\beta}+n}$, gives us the posterior mean for the examiner. Intuitively, this procedure shrinks an examiner's NPE share towards the mean NPE share in the art unit, more so when the examiner has granted few patents.

Finally, we compute deviations from art unit prior mean for each examiner in a given year. We average these shrunk examiner effects across years, in order to make them comparable to the Kane Staiger shrunk examiner effects. The distribution of shrunk examiner effects computed using this procedure are shown on Figure 3.3. This distribution looks very similar to Figure 3.2, with a standard deviation around 44.4% of the baseline and an inter-quartile range of 8.8%, as opposed to the 24.0% and 6.9% reported in Table 3.6. The main difference is that the distribution computed using the Empirical Bayes Count Model has thicker tails. The fact that the two distributions have similar characteristics is a re-assuring fact, since the Empirical Bayes Beta-Binomial count model imposes much weaker parametric restrictions on the prior distribution.[32] This suggests that we can treat the results from the more tractable Kane-Staiger methodology as an approximation to a more flexible model of examiner effects.

Figure 3.3.: Distribution of NPE Examiner Effects (Empirical Bayes Count Model)



Notes: A kernel density plot of shrunk examiner effects computed using the Empirical Bayes Count Model described in Section 3.4.4.2. The figure is re-scaled to match Figure 3.2, and examiner effects above 0.01 and below -0.01 are moved to 0.01 and -0.01, respectively. The examiner effects are weighted by cases examined.

The results presented in Figure 3.3 still rely on an untestable parametric assumption about the prior, which we have assumed to be a Beta distribution. A more flexible prior would

[32]In fact, we can compare the consistency of the examiner rankings across the two approaches. We show evidence of this in Appendix D.

involve a mixture of Beta distributions, but the downside of such an approach is having to estimate additional hyperparameters with only twenty or so data points on examiners in a given art unit and year.

3.5. Distinguishing Between Mechanisms

In this section, we present a series of results suggesting that the large spread in examiner NPE effects, which we documented in Section 3, can be linked to certain examiner tendencies observable in our data. We will investigate whether causal examiner effects (tendencies) based on examination actions are predictive of NPE purchase and litigation. We generally find that examiner tendency to use 103(a) and 112(b) blocking actions is *negatively* predictive of NPE outcomes. In addition, we find evidence that this operates through changes to patent claims language between initial application and final grant. We also run a series of predictions for other important patent market outcomes, and draw comparisons to the NPE findings.

3.5.1. Predictive Regressions Methodology

Our core methodology in this section is to predict outcomes such as NPE purchase using causal examiner effects, which are computed using the methodology from the previous section. One way of viewing our approach is that we want to predict outcomes using variation in randomly assigned co-producer characteristics.

Formally, we compute causal examiner effects leaving out data from the application we are trying to predict. This purges the regression of any mechanical correlations based on the nature of the patent itself affecting examiner behavior measures. We then set up the following regression at the patent level:

$$NPE_{ij} = \beta \hat{E}_j + \epsilon_{ij}$$

where i indexes the patent and j indexes the examiner. $\hat{E}$ represents some vector of computed leave-one-out examiner effects, such as propensity to use 103(a) blocking actions or to force changes to claims text. This is a simple way of documenting which examiner behaviors correlate with NPE purchase.

We first run predictive regression with single examiner effects. The advantage of this regression is that it allows us to analyze each effect in isolation. Next, we also run some "horse race" regressions, including multiple examiner effect measures. Once put in a horse-race with other variables such as allowance rates and changes in patent claims text, the blocking action variables tend to lose importance. Our interpretation of this is that various blocking actions force different types of changes to the patent text, but our text-based measures are too crude to distinguish between these types of changes. Therefore, it makes more sense to trace the changes back to the source, and focus on the blocking actions for interpretability. We also run horse-race regressions using only blocking action examiner effects.

From a methodological perspective, the leave-one-out examiner effect should ideally be uncorrelated with the error term ϵ_{ij}, which captures unobserved characteristics of the patent. Note that selection effects might introduce a bias - the pool of granted patents allowed by each examiner varies by examiner characteristics. For example, if examiners with a high rate of 103(a) blocking actions have higher quality patents on average, and NPEs purchase lower quality patents, then ϵ will be negatively correlated with $\hat{E}_{103a}$, and our estimated β would overstating the intensive margin effect of the blocking action. We check for this by including additional controls based on inventor experience and application characteristics (similar to the "Additional Patent Control" specification in Table 3.3), which has negligible impact on the estimated coefficients, and changes in coefficients move both ways.

An obvious alternative method would be to take the set of tendencies for each examiner, and correlate the tendencies at the examiner level (e.g. correlate examiner NPE effect with examiner 103(a) effect). We do run this exercise by splitting the sample to avoid picking up mechanical correlations, and find similar qualitative results.

3.5.2. Results

We run pairwise regressions predicting NPE purchase using various causal examiner effects, including propensity to change the number of words per claim, propensity to remove independent claims, and propensity to use various blocking actions discussed in Section 3.2.1. In the results reported in the main body, we use blocking actions on eventually granted patents, to better capture intensive margin effects, but we also report results for blocking action usage at the application level in the Appendix.

We compute how much a one-standard-deviation change in the examiner causal effect would change the outcome in question (e.g. NPE purchase), as a fraction of the baseline outcome rate. These results are reported in the tables in this section. In addition, we present summary graphs showing the pairwise predictive power of various examiner effects. In the graphs, we normalize the computed effects by the "own variable" predictive effect. This own variable effect (e.g. the examiner NPE effect) will generally have the strongest predictive power, since it's constructed to be predictive of the outcome, but this result is not mechanical and does not have to hold. Finally, we color the bars blue for variables that are positively predictive of the outcome and red for the variables that are negatively predictive. The p-value for the coefficients are displayed in white.

3.5.2.1. Purchased Patents

As we see in Table 3.9 and Figure 3.4, the correlation between NPE outcome and examiner NPE effect is positive and the strongest predictor, with a one standard deviation change in examiner effect translating to a 42.5% change relative to the NPE purchasing baseline. For the other examiner effects, allowance rate is also positively predictive, with a magnitude around a quarter of the baseline effect size. Beyond that, there are several examiner effects that are *negatively* predictive of NPE purchase. These include examiner tendencies to add more words to claims, issue blocking actions based on failure to satisfy non-obviousness (103(a)) and claim clarity (112(b)). Each of these have a sizable amount of predictive power relative to the NPE examiner effect, ranging from 5-15% of the baseline rate. This is more easily seen in Figure 3.4, which reports the same results, but ordered by the relative magnitude of the effect.

Based on the results, NPEs appear to purchase patents that have fewer words added during examination than the average patent, which is a sign of less specificity or clarity. The examiner usage of blocking actions allows us to further interpret the nature of these word changes. Examiners who use more 112(b) blocking actions are less likely to have their patents purchased by NPEs (a 5-8% effect per examiner effect standard deviation), although this result is not as strong in terms of magnitude and statistical significance as some of the other effects we analyze. In combination with the evidence on word counts, this provides indirect evidence that examiners who use more 112(b) actions force more clarifications to patent claims during the examination process. In addition, examiners who use fewer 103(a) blocking actions are also more likely to have their patents purchased by NPEs (a 10% effect per examiner effect distribution), suggesting that they do not force applicants to specify the non-obvious aspects of the invention in question.[33] Finally, there is also a borderline significant effect on the change in the number of independent claims during examination, which suggests that NPEs prefer patents that have more independent claims. This property has been discussed in the legal and economic literature as a proxy for the strength of the patent, as patents with more independent claims are less likely to be completely invalidated. The extra claims may also represent broader patent scope, essentially covering more of

[33] Various legal guides suggest that one way to respond to a 103(a) blocking action is to add enough specifics so that the invention is no longer implied or suggested by any prior art.

intellectual property space, and allowing NPEs to find more possible infringing products.

As a comparison, we run the same routine, except for patents purchased by firms not on our NPE list. This exercise yields additional insights into NPE behavior, as it is possible that they behave just like regular firms in the markets for technology. The results are reported in columns 3 and 4 of Table 3.9, and suggest that regular firms purchase different types of patents relative to NPEs. In general, the estimated importance of various examiner tendencies is much smaller in magnitude for non-NPE purchases relative to NPE purchases. On the specific effects, 103(a) and 112(b) usage are much smaller in magnitude (1.3% and 0%, respectively).[34] Later in the section, we formally confirm that there are significant differences by running conditional regressions.

[34]As an aside, the 112(a) examiner effect appears to be unrelated to non-NPE purchase, suggesting that addition clarity in technological disclosure does not facilitate more transfer through purchases. It would be interesting to see if this holds in the pool of licensed patents.

Table 3.9.: Pairwise Examiner Effects Results - Purchase

Panel A: General Examiner Effects

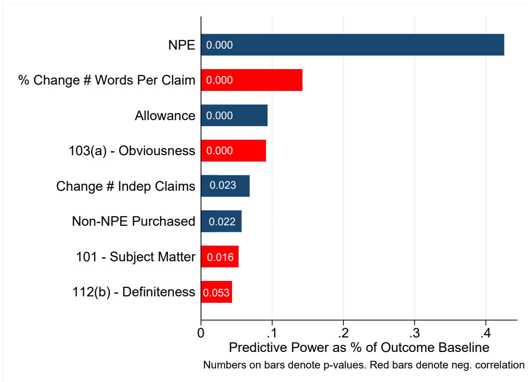
	NPE Purchase	NPE Purchase	Non-NPE Purchase	Non-NPE Purchase
NPE Purchase	0.425***	0.447***	0.015**	0.014*
	(0.066)	(0.076)	(0.005)	(0.005)
Non-NPE Purchase	0.057**	0.044^{+}	0.112***	0.110***
	(0.022)	(0.025)	(0.005)	(0.006)
Words Per Claim Change	-0.142***	-0.148***	-0.018***	-0.019***
	(0.021)	(0.021)	(0.004)	(0.004)
Independent Claims Change	0.056^{+}	0.078**	-0.000	0.005
	(0.029)	(0.030)	(0.006)	(0.006)
Patent-Level Controls		x		x
N	1,269,623	1,269,623	1,269,623	1,269,623

Panel B: Examiner Blocking Action Effects

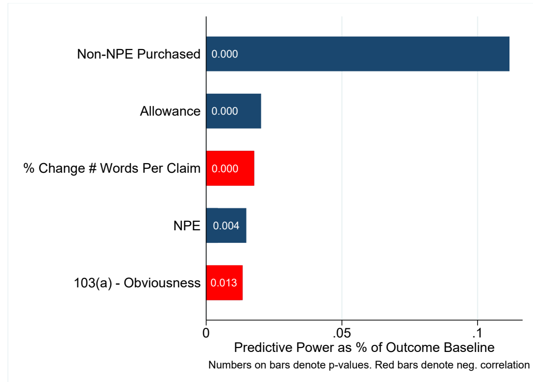
	NPE Purchase	NPE Purchase	Non-NPE Purchase	Non-NPE Purchase
101	-0.052*	-0.064**	0.003	0.002
	(0.022)	(0.024)	(0.003)	(0.003)
102(a)	0.018	0.012	0.005	0.006
	(0.024)	(0.025)	(0.005)	(0.005)
103(a)	-0.091***	-0.099***	-0.013*	-0.013*
	(0.023)	(0.023)	(0.005)	(0.005)
112(a)	-0.020	-0.021	0.003	0.003
	(0.017)	(0.018)	(0.004)	(0.004)
112(b)	-0.043^{+}	-0.047*	-0.000	0.001
	(0.022)	(0.023)	(0.004)	(0.004)
Patent-Level Controls		x		x
N	1,269,623	1,269,623	1,269,623	1,269,623

Notes: All results are produced using the restricted Frakes and Wasserman data. All coefficients are normalized to represent the impact of a one standard deviation in examiner effect relative to the baseline rate of the outcome in question (1.0% for NPE patents, 19.7% for non-NPE purchased patents). Words per claim variable refers to the percentage change in word count between application and grant. Patent-Level Controls include the history of the first inventor and the log of the number of independent claims at application. Standard errors are clustered at the examiner level. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

Figure 3.4.: Graphical representation of the mechanism results for pairwise prediction of patent purchases



(a) Predicting NPE Patent Purchase



(b) Predicting non-NPE patent purchase

Notes: Results presented are based on the pairwise predictive regressions reported in Table 3.9. Effect sizes are again calculated in terms of fraction relative to the baseline rate of NPE purchase per one standard deviation change in the examiner effect, and then normalized to the size of the examiner effect calculated from the outcome variable in question (NPE purchase, non-NPE purchase). Blue bars indicate a positive predictive coefficient and red bars indicated a negative predictive coefficient.

3.5.2.2. Litigated Patents

Next, we move onto the issue of litigation behavior. We can essentially repeat the same exercise as before, except with NPE and non-NPE litigated as the outcomes. We report the corresponding results in Table 3.10. One challenge here is that the baseline rate of non-continuation, NPE-litigated patents is around 0.033% of the patent population (about one-tenth the rate of non-NPE litigated patents).[35] Our methodology for computing examiner effects not work well with very rare outcomes, as it relies on consistency to extract the underlying signal. Therefore, while we do compute an examiner "NPE Litigated" effect, it is quite noisy and not a reliable predictor of outcomes.

[35] One important issue to note here is that NPEs often use the same set of patents against many plaintiffs, much more so than regular firms. Therefore, despite having about a tenth of the number of asserted patents, NPEs are still a major presence in terms of the number of firms they sue for infringement. In the Cotropia et al classifications of 2010 and 2012 IP lawsuits, around 35% of plaintiffs are NPEs. These numbers may also be a little misleading, as pre-AIA joinder rules meant NPEs could target multiple defendants in the same case, which we observe in the litigation data provided by RPX.

The results suggest that NPEs litigate patents that have some similarity to patents litigated by regular firms, based on the strong relationship between NPE litigated patents and non-NPE litigated examiner effect. However, NPEs once again appear to have a stronger sensitivity to other examiner effects, such as 103(a) and 112(b), although non-NPE litigated patents do show sensitivity to the same two examiner effects. These results could also be interpreted through the lens of standard litigation models, which suggest that plaintiffs and defendants only go to court if there is disagreement over the chances of winning. Vague claims language and uncertainty over obviousness may create such disagreements. Both types of litigated patents also contain more independent claims than they would have if examined by the average examiner. Finally, the results also exhibit strong 112(a) effects, which did not show up for NPE purchasing and is harder to interpret.

Table 3.10.: Pairwise Examiner Effects Results - Litigation

Panel A: General Examiner Effects

	NPE Litigated	NPE Litigated	Non-NPE Litigated	Non-NPE Litigated
NPE Litigated	-0.120	-0.119	-0.070**	-0.055*
	(0.140)	(0.132)	(0.025)	(0.023)
Non-NPE Litigated	0.238**	0.157*	0.392***	0.267***
	(0.088)	(0.061)	(0.082)	(0.040)
Words Per Claim Change	-0.350***	-0.346***	-0.070***	-0.061***
	(0.057)	(0.046)	(0.016)	(0.016)
Independent Claims Change	0.117*	0.130*	0.050**	0.049**
	(0.052)	(0.054)	(0.019)	(0.019)
Patent-Level Controls		x		x
N	1,269,623	1,269,623	1,269,623	1,269,623

Panel B: Examiner Blocking Action Effects

	NPE Litigated	NPE Litigated	Non-NPE Litigated	Non-NPE Litigated
101	-0.135*	-0.123^{+}	-0.043*	-0.032*
	(0.063)	(0.066)	(0.018)	(0.015)
102(a)	0.004	0.011	-0.013	-0.011
	(0.050)	(0.054)	(0.017)	(0.016)
103(a)	-0.236***	-0.240***	-0.062**	-0.039*
	(0.062)	(0.066)	(0.020)	(0.017)
112(a)	-0.139**	-0.140**	-0.046^{+}	-0.036*
	(0.043)	(0.044)	(0.027)	(0.017)
112(b)	-0.117*	-0.139*	-0.038^{+}	-0.040*
	(0.055)	(0.055)	(0.020)	(0.018)
Patent-Level Controls		x		x
N	1,269,623	1,269,623	1,269,623	1,269,623

Notes: Results are run on the restricted Frakes and Wasserman sample. All coefficients are normalized to represent the impact of a one standard deviation in examiner effect relative to the baseline rate of the outcome in question (0.5% for non-NPE litigated and 0.033% for NPE litigated).

3.5.2.3. Testing for Differences

Finally, we look for examiner effects that distinguish between NPE purchase vs. non-NPE purchase, NPE litigated vs. NPE purchase, and NPE vs. non-NPE litigated. As an additional exercise, we test for a distinction between Intellectual Ventures and other NPEs. So far, we have compared groups of patents with interesting outcomes to the rest of the patent pool. Here, we restrict the analysis sample to patents with at least one of a pair of outcomes, and then looking for examiner effects that are conditionally predictive of the first effect in each pair mentioned. This gives us a more rigorous way to compare types of patent outcomes.

The results are shown in Table 3.11. NPEs purchase patents from examiners who use 103(a) and 112(b) blocking actions less often than their peers, and strongly prefer patents with fewer added words during examination. NPEs selectively litigate more obvious patents, but there is no significant additional 112(b) effect. One other interesting point to note is that the patents purchased by Intellectual Ventures do not exhibit significant differences relative to other NPE purchased patents, as shown by the several precisely estimated zero results in column 4. This provides some evidence that the purchasing activities of Intellectual Ventures, which ignores their in-house inventions, are not particularly different from other NPEs, contrary to narratives that distinguish types of NPEs.

Table 3.11.: Pairwise Differential Outcome Regressions

	NPE vs. non-NPE		Litigated vs. Purchased	Int. Vent. vs. Other
	Purchase	Litigated	NPE	Purchase
101	-0.042**	-0.073	-0.046	-0.012
	(0.015)	(0.050)	(0.033)	(0.010)
102(a)	0.011	0.018	-0.012	-0.005
	(0.020)	(0.056)	(0.044)	(0.016)
103(a)	-0.071***	-0.190**	-0.123*	-0.004
	(0.020)	(0.067)	(0.049)	(0.013)
112(a)	-0.019	-0.074+	-0.133**	0.026*
	(0.014)	(0.042)	(0.045)	(0.012)
112(b)	-0.039*	-0.074	-0.066	-0.000
	(0.019)	(0.053)	(0.046)	(0.014)
Words Per Claim Change	-0.114***	-0.298***	-0.178***	-0.016
	(0.017)	(0.060)	(0.043)	(0.012)
Independent Claims Change	0.066*	0.080	0.056	-0.003
	(0.028)	(0.065)	(0.055)	(0.028)
N	262,511	7,156	12,953	12,953

Notes: The results are is run on the restricted Frakes and Wasserman data. Int. Vent. refers to Intellectual Ventures. All coefficients are normalized to represent the impact of a one standard deviation in examiner effect relative to the baseline rate of the outcome in question. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

3.5.2.4. Discussion

Overall. the general pattern in the mechanism results show that patents with fewer words added during examination and fewer independent claims removed are more likely to show up in all of the patent groups we have analyzed. Digging further down, NPE purchased and NPE litigated patents have much more in common with non-NPE litigated patents than with non-NPE purchased patents. NPEs tend to purchase and litigate patents granted by examiners who are less likely to judge a patent application to have vague claims and to be obvious. This holds up in comparisons to both the general patent pool and to corresponding non-NPE purchased and litigated patents.

Our results are inconsistent with the theory that NPEs serve as effective intermediaries in identifying valuable technologies in the pile of vaguely worded patents.[36] If this were the case, then NPEs would be playing an efficient screening role, dealing with the problems introduced into the IP system through an inconsistent examination process. There are two issues with this critique. First, even if NPEs play a role in screening vague patents, the current setup involving high legal fees seems like a very costly way to deal with the problem. Our policy calculations in Section 3.6 suggest that it would be much more cost-effective to solve the vague patents problem at its source, namely the patent examination process. Second, the activity of NPEs appear to be at best unrelated if not negatively related to technological attributes of the patents. NPEs purchase patents from examiners who issue more obvious patents, which suggests that these patents are technologically less innovative relative to others in their cohort. Finally, NPE purchased patents are unassociated with 112(a), the technological disclosure requirement. This suggests that NPEs are better at uncovering legally useful patents rather than technologically useful patents.

3.5.3. Robustness Checks

The main focus in this section will be on testing robustness to NPE classification, adding extra controls to the basic specification, and testing for the significance of examiner career effects. We report the most important results in Tables 3.12 and 3.13 and leave the rest for the appendix.

3.5.3.1. Alternative Specifications

First, we verify that our results are robust to changes in NPE classification. For this, we construct a portfolio from the Cortropia et al. classifications, focusing on entities they classify as large aggregators and small patent holding companies. As mentioned earlier, their list

[36] The idea of digging up valuable patents is discussed in the book “Rembrandts in the Attic” by Kevin Rivette and David Kline.

only identifies entities that are plaintiffs in litigation in 2010 and 2012, and therefore may leave some true NPEs in the control group. The results using their list of NPEs are consistent with our core findings.

We also run alternative specifications to compute the residuals in our examination effects methodology, which corresponds to changing the X in Equation **??**. The baseline just contains art unit by year fixed effects as controls, to reflect the level of random assignment. One major addition we try here is to add assignee fixed effects, based on the initial assignee on a given patent.

The limitation of this specification is that it eliminates many unassigned patents, which make up a sizable part of NPE portfolios, as shown in Table 3.4. However, this approach allows us to compare the patents sold by a firm to NPEs versus the ones they keep or sell to other entities. The results suggest that firms tend to sell the vaguer and more obvious patents within their own portfolios.[37] An additional interpretation of these results is that it rules out the story that our results are driven solely through a bankruptcy channel, namely the possibility that NPEs mostly purchase intellectual property assets during firm liquidation and firms that own the types of patents shown in our results are more likely to go bankrupt. Our results, as shown in Table 3.12, are generally very similar to the baseline results, but the additional noise from a reduced sample pushes the 112(b) result to become statistically insignificant.

We also run specifications adding in only observable patent application characteristics as controls. These controls include the number of independent claims at application, the applicant entity size (regular or small), and the patenting history of the firm and the inventors at the time of the application. The results are also robust to these additions.

[37]This result may be related to the concept of "patent privateering." Practicing companies face the threat of countersuit if they assert their patents directly against a competitor. Instead, they can choose to go after competitors by selling their patents to NPEs with some sort of protection provision included in the transaction.

Table 3.12.: Robustness checks for key predictors NPE purchase.

Specification	103(a)	112(b)
Baseline	-0.091***	-0.043^{+}
	(0.023)	(0.022)
Kesan *et al* NPEs	-0.090***	-0.051*
	(0.027)	(0.023)
Non-IV NPEs	-0.088***	-0.042^{+}
	(0.027)	(0.026)
IV	-0.095**	-0.044^{+}
	(0.030)	(0.026)
Assignee Fixed Effects	-0.096***	-0.037
	(0.025)	(0.023)
Additional Patent Controls	-0.088***	-0.042^{+}
	(0.026)	(0.025)
Five Year Lag	-0.092***	-0.044^{+}
	(0.025)	(0.023)

Notes: Non-IV NPEs refers to patents purchased by NPEs other than Intellectual Ventures and IV refers to patents purchased by Intellectual Ventures. Additional Patent Controls refers to adding the same inventor and application characteristics to the basic estimating equation as the specification in Table 3.8. Five Year Lag refers to computing signal covariances using data five years apart instead of one, and is the same methodology used to compute the results in Table 3.13. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

3.5.3.2. Examiner Career Effects

One possible concern is that examiner effects are not "fixed" as in our preferred framework but rather vary substantially over the course of an examiner's career at the patent office. For instance, Frakes and Wasserman (2014) find that time constraints vary a lot over the course of an examiner's career. Our baseline methodology does not account for this, as we assume a fixed examiner effect μ_j, along with an idiosyncratic examiner by cohort effect θ_{jt}.

We show that time-varying examiner effects are not quantitatively important in a number of ways. First, we re-compute results by computing signal covariances using different lags $\hat{\sigma}_{\mu}^2 = cov(\bar{v}_{jt}, \bar{v}_{j(t+k)})$, for values of k beyond 1, which we use to generate our baseline results. As shown in Table 3.13, our core results are robust to these changes. We compute NPE effect distribution parameters for $k = 3, 5$, and compare it to the parameters estimated by using $k = 1$, but restricting the samples to examiners with at least 4 or 6 years of experience, respectively. Here, we use the non-continuation PERD sample, in order to increase the years of coverage, which is not very large relative to the lags we are using. The results suggest that our high signal covariance is not driven by examiner career effects, as we are now correlating residuals for examiners at very different points in their respective careers. The decrease in estimated signal correlations might not be driven by career effects alone, as time variation in NPE purchase rates could cause problems in our thirteen-year sample.

As a methodological aside, relative to the non-continuation PERD sample results reported in Table 3.8, the modified baselines we compute here have almost identical distribution parameters, suggesting that our methodology shrinks examiners with less data in an effective manner.

Table 3.13.: Robustness checks for the effect of Examiner Career Effects on computed NPE effect distribution.

Specification	Signal SD	Shrunk Effect SD	Shrunk Effect IQR
Three Year Lag	44.3%	19.4%	4.8%
Baseline	53.2%	26.7%	6.4%
Five Year Lag	35.8%	14.5%	3.5%
Baseline	52.8%	27.8%	6.1%

Notes: Three Year Lag and Five Year Lag Refer to computing the signal covariance $\hat{\sigma}_{\mu}^2 = cov(\bar{v}_{jt}, \bar{v}_{j(t+k)})$ using $k = 3, 5$ respectively. The baseline in each case is re-calculated to only include examiners that have careers of at least 4 and 6 years, respectively. The results reported here are computed using the full PERD sample.

Second, we follow the methodology in Frakes and Wasserman (2015), and look for significant changes in examiner NPE grant rates over their careers, particularly before and after promotions. With minor exceptions, we generally do not find significant NPE effect differences over an examiner's career. These results are available from the authors upon request.

3.5.3.3. Results Breakdown

First, we break down the results by Intellectual Ventures versus other NPEs in our list. A downside to our methodology, as discussed in the context of computing NPE Litigated examiner effects, is that it has difficulty extracting signals that occur with very low frequency. Therefore, we cannot use it to describe the behavior of NPEs that hold a handful of patents. However, we are still able to check the behavior of Intellectual Ventures, which has by far the largest patent portfolio, versus the combined portfolio of other NPEs.

In addition, we also break down our resultsby technological areas. We do this by analyzing patents from each of the eight technology centers at the USPTO separately.[38] Results are reported in Table C.13. Our results suggest that the non-obviousness result holds mainly for the four IT-related technology centers - 21 (Computer Architecture, Software, and Information Security), 24 (Computer Networks, Multiplex communication, Video Distribution, and Security), 26 (Communications), and 28 (Semiconductors, Electrical and Optical Systems and Components) - and the vague claims text results hold for art unit 26 and 28. Our core results do not hold for technology centers 16 (Biotechnology and Organic Chemistry), 17 (Chemical and Materials Engineering), and 36 (Transportation,Construction, Electronic Commerce, Agriculture, National Security and License & Review), suggesting that the IP market operates differently in those areas. The results for the non-NPE litigated patents generally follow the same patterns.

3.5.3.4. Horse Race Between Effects

Finally, we also run a "horse race" regression between the examiner effects to predict various outcomes. The results are shown in Table 3.14. The 103(a) examiner effect tends to show through in all of the specifications. One issue to note is that some pairs of examiner effects

[38] There are a few dozen art units within each technology center.

have low correlation, and therefore each separate effect has some predictive power, particularly the 101 effect. However, as shown in Table C.1, 103(a) and 112(b) tend to be used together on applications, making it harder to identify the contribution of the 112(b) effect.

Table 3.14.: Horse Race Regressions Predicting Each Outcome

Panel A: Horse Race for Outcomes

	NPE Purchase	Non-NPE Purchased	NPE Litigated	Non-NPE Litigated
101	-0.0425^{+}	0.0041	-0.0977	-0.0314
	(0.0225)	(0.0042)	(0.0636)	(0.0185)
102(a)	0.0313	0.0053	0.0472	0.0000
	(0.0241)	(0.0051)	(0.0518)	(0.0184)
103(a)	-0.0829***	-0.0166**	-0.202**	-0.0472*
	(0.0249)	(0.0058)	(0.0635)	(0.0201)
112(b)	-0.0104	0.0014	0.0241	0.00428
	(0.0241)	(0.0054)	(0.0593)	(0.0249)
N	1,269,623	1,269,623	1,269,623	1,269,623

Panel B: Differential Horse Race Regressions

	NPE vs. non-NPE		Litigated vs. Purchased
	Purchase	Litigated	NPE
101	-0.0327*	-0.0427	-0.0226
	(0.0163)	(0.0521)	(0.0340)
102(a)	0.0223	0.0406	0.00962
	(0.0198)	(0.0567)	(0.0439)
103(a)	-0.0607**	-0.170*	-0.0989^{+}
	(0.0212)	(0.0715)	(0.0508)
112(b)	-0.0124	0.0152	0.0449
	(0.0209)	(0.0583)	(0.0549)
N	262,511	7,156	12,953

Notes: The regressions in Panel A are run on the baseline sample. The regressions in Panel B are run on the same restricted samples as in Table 3.11. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

3.5.3.5. Direct Evidence on Weak Patents

In ongoing work, we relate the patent acquisition behavior of NPEs more directly to the validity of the patents by studying whether the examiners who produce NPE patents tend to be invalidated in court (in the case of litigation by regular entities) or to be reversed during appeal procedures at the Patent Trial and Appeal Board (in the case of an appeal by a regular entity during the course of prosecution). Preliminary results suggest that the examiners with a high NPE effect are more likely to behave in a way that is not in accordance with the law. Figure C.7, reported in Appendix B, shows that examiner with a high NPE effect tend to be reversed at the Patent Trial and Appeal Board when there is an appeal during the course of prosecution.[39]

3.6. Policy Implications

3.6.1. Calibrations

In this section, we use our computed examiner effects in order to estimate the possible returns to investment in the patent examination process. We focus here on the possible litigation fees saved from a reduction in the types of patents currently used by NPEs, and consider this to be a lowerbound on the social costs of issuing such patents. Our computation does not include other possible inefficiencies generated by NPEs, such as implicit taxes imposed on products challenged by NPEs (Tucker (2015)) and other possible distortions in the intellectual property system.

First, we discuss here some estimates for litigation and USPTO costs that we will use to interpret our calculations. In terms of annual litigation fees associated with NPE lawsuits, there are a variety of estimates in the legal literature, each using a different methodology. The main sources for this data comes from the American Intellectual Property Law Association (AIPLA) annual reports, which surveys its members for estimates of how much they are paid

[39]More than 50% of examiners have one of the patent applications reviewed by the Patent Trial and Appeal Board during the course of their career, which mitigated concerns about the selection effects inherent in our strategy.

for cases of different sizes, and the surveys conducted by RPX and Bessen and Meurer (2014), which surveys defendants to estimate spending on legal defense and other costs. A major difference in the figures reported by the two sources stems from the fact that the AIPLA reports median figures for each case size category, whereas Bessen and Meurer use means, which are much higher because of a long right tail. The figure arrived at by Bessen and Meurer is $29 billion in "direct costs," of which around 17% come from defensive litigation costs.[40] Therefore, this comes out to over $5 billion used in defense. Although this number may be on the high side, it also does not include resources used by NPEs, so we use this number for our calculations. In terms of USPTO spending, the patent portion of the annual USPTO budget is $3.13 billion for FY 2016.[41] Of this, around $800 million is paid out in examiner salary.[42]

The cost of policy initiatives about examiners is typically modest in comparison to the potential benefits mentioned above. For instance, the aforementioned *Enhanced Patent Quality Initiative* (EPQI) at USPTO will cost around $11 million for FY 2016, ramping up to about $31 million by FY 2020.[43]. Another quality initiative implemented by the USPTO was the "second pair of eyes" review,[44] which was first piloted on the set of business method patents (USPTO class 705), then partially expanded in 2003 to several other technology centers and eventually discontinued in the late 2000s.[45] The initiative introduced one hour of review per provisionally granted patent, in order to flag obvious issues. Given that examiners spend on

[40] The remainder is licensing fees, which we interpret in a static way as just a transfer, and therefore not an inefficiency.

[41] http://www.uspto.gov/sites/default/files/documents/fy16pbr.pdf

[42] Based on examiner grade-level data from Frakes and Wasserman plus data on salary by examiner grade available in the USPTO annual budget report.

[43] Based on P55 of the 2016 USPTO Budget Report, available at http://www.uspto.gov/sites/default/files/documents/fy16pbr.pdf

[44] Discussed on P28 of the 2006 USPTO Budget Report, available at http://www.uspto.gov/sites/default/files/web/offices/ac/comp/budg/fy07pbr.pdf

[45] Based on discussions with USPTO policy directors, the discontinuation resulted from examiners being uncomfortable with this review process, and possibly granting fewer patents in response, as only granted patents were scrutinized.

average 19 hours reviewing each application, this suggests that such a policy, if implemented across all technology centers, would cost at most \$40 million.[46]

As mentioned above, the estimated NPE litigated examiner effects are quite noisy, due to the sparsity of the outcome combined with the nature of our methodology. Instead, we can use a series of alternative approaches to come up with an estimate of the partial equilibrium cost savings from the two aforementioned policies. The results will be partly based on results reported in Table 3.9. We can approximate the effect of policies like the EPQI, which aim at establishing best examination practices, as moving examiners above the 75th percentile in NPE purchase effect to the 75th percentile, or moving examiners from below the 25th percentile in 103(a) and 112(b) usage to the 25th percentile. Then, we can compute the impact of such a change as a fraction of the standard deviation of the examiner effect distribution, and then scale it by the coefficient in our mechanism result.

The calculations are reported in Table 3.15. The first row shows the returns if EPQI can directly target the NPE purchase examiner effects. The \$297.5 million estimate suggests a very high social return on the \$11 million investment. The remaining rows show the effects of EPQI if it were to target examiner usage of blocking actions. Both calculations also yield significant social returns.

[46] Taking the \$800 million as a base number and dividing it by the hours ratio. This will be an overestimate, as the there are more applications than granted patents.

Table 3.15.: Simulating the Benefits of the Enhanced Patent Quality Initiative

	ΔNPE Litigation per SD	Move tail (SDs)	Total ($ millions)
NPE Purchase	0.238	6%/24% =0.25	297.5
103(a)	0.091	2.8%/19.1%=0.146	66.7
112(b)	0.043	7.7%/38.6%=0.199	42.8

Notes: Results are calculated by taking the baseline number of $5 billion per year in NPE litigation fees, multiplying it by the effect per SD, and then multiplying by the effect of the policy (represented in standard deviation units). The second column is based on results reported in Table 3.9. The third column, "Move tail," reflects the change by moving examiners with NPE Purchase effects above the 75th percentile to the 75th percentile and by moving examiners with 103(a) and 112(b) usage below the 25th percentile to the 25th percentile, similar to the numbers reported in Table 3.6, but then re-scaled to be a fraction of the standard deviation of the examiner effects distribution for that variable.

Next, we can evaluate the second-pair-of-eyes policy using a slightly more involved simulation. Here, we once again rely on translating other examiner effects into a change in NPE litigation. We simulate such a policy using our estimated primitives by randomly assigning each application i, which was assigned to examiner j, to another examiner k, and taking the minimum of the examiner NPE effects or the maximum of the blocking action effects. We winsorize examiner effects at the 1st and 99th percentile before performing the calculations:

$$\Delta NPE = \frac{1}{N}\sum_{i=1}^{N}\min(v_j, v_k) \tag{3.2}$$

$$\Delta 103(a) = \frac{1}{N}\sum_{i=1}^{N}\max(v_j, v_k) \tag{3.3}$$

In addition, we also simulate a scenario in which the second examiner has a much small effect, which is captured in Equation 3.4:

$$\Delta NPE = \frac{1}{N}\sum_{i=1}^{N}\min(v_j, \frac{2}{3}v_j + \frac{1}{3}v_k) \tag{3.4}$$

$$\Delta 103(a) = \frac{1}{N}\sum_{i=1}^{N}\max(v_j, \frac{2}{3}v_j + \frac{1}{3}v_k) \tag{3.5}$$

Table 3.16.: Simulating the Benefits of the "Second-pair-of-eyes" Policy

	Effect per SD	Major Second (SDs)	Total ($ millions)	Minor Second (SDs)	Total ($ millions)
NPE Purchase	0.238	13.3%/24%=0.554	659.4	4.23%/24%=0.176	209.7
103(a)	0.091	12.5%/19.1%=0.654	297.8	4.40%/19.1%=0.230	104.8
112(b)	0.043	22.9%/38.6%=0.593	127.6	7.90%/38.6%=0.205	44.0

Notes: Results are calculated by taking the baseline number of $5 billion per year in NPE litigation fees, multiplying it by the effect per SD, and then rescaling by the impact of the policy, measured in standard deviation units. The second column is based on results reported in Table 3.9. "Major Second" refers to the scenario in which the second examiner has a major impact on the examination process (see Equation 3.2), whereas "Minor Second" refers to the scenario in which the second examiner has a minor impact (see Equation 3.4).

3.6.2. Discussion

In general, our calibration results suggest large social returns to relatively the inexpensive USPTO investments. Of course, our calculations do not take into account the reactions of NPEs to a policy change. In addition, our calculations also equally weigh litigated patents, when in reality, some patents are involved in a disproportionate number of cases. The nature of our methodology somewhat limits our ability to accurately account for these weighting issues. Finally, our calculations have several moving pieces, particularly the modeling of the two policy initiatives. However, we have tended to err on the cautious side in all of our assumptions, in order to provide a lowerbound for the returns to public investment. More broadly, our evidence suggests that policy reforms about patent examiners have great potential and would be a welcome addition to the current policy debate, where the discussion has focused on reforms of the statutes in Title 3 of the US code or on reforms of the court system.

Various limitations of this paper could be addressed in future work and would help refine the policy implications. First, we abstract away from the market for patents in our current analysis. A key issue when thinking about the patent market and NPEs is to determine why practicing companies do not outbid NPEs for these patents, or at least drive up the price. One simple explanation is that NPEs have greater expertise in identifying patents useful for their business model. The patent market is not particularly liquid, and some NPEs have built up extensive patent agent networks (Hagiu 2011). Another explanation is that regular companies cannot use patents in the same way as NPEs, because they are limited by the threat of a countersuit and possibly damages to firm brand, and therefore NPEs have much higher valuations for offensive patents than regular firms. Finally, practicing firms suffer from a free riding problem, because their purchasing of a patent for defensive purposes also helps their rivals in the same market. A recent trend in the market is the emergence of defensive aggregators, such as RPX and Allied Security Trust. These defensive aggregators are funded by industry groups, and may offer a partial, market-based solution to the NPE problem by buying up problematic patents and guaranteeing that they will not assert them. In addition, NPEs may still be somewhat active even if the pool of patents possessed fewer of the attributes they currently appear to desire.

Second, we have not delved into the details of the court system. An important caveat here is that the effectiveness of patents in litigation depends not only on the claims text, but also on how judges interpret the claims. A patent with conspicuously broad or obvious claims would therefore be less useful to NPEs looking to maximize litigation revenue or the threat of litigation. An obvious channel for leveraging these patents, as frequently discussed in NPE debates, is the Eastern District of Texas, which has had a reputation for favorable treatment of plaintiffs. Another problem in the analysis is that we do not observe legal fees and settlements, and therefore miss out on heterogeneity across lawsuits.

Finally, we did not attempt to fully evaluate the welfare effects of the activities of NPEs. Our results only address issues of static efficiency, and it may be that NPEs bolster exit value in a way that encourage entry of small firms or individual inventors, pushing against under-investment in innovation activities. However, our results suggest that NPEs would be encouraging the creation of vague and obvious patents, which seems like a blunt instrument for increasing the expected value of entry. Policies such as expanded R&D tax credits to small firms may be much more efficient socially to achieve such goals.

3.7. Conclusion

In this paper, we have exploited the random assignment of patent applications to examiners to show that examiners have a significant influence on a variety of patent outcomes, and to shed light on the characteristics of patents purchased and litigated by non-practicing entities. The evidence suggests that NPEs purchase more patents that are more obvious and contain vaguer claims text, attributes unrelated to or negatively associated with social value and invention quality. This research offers both evidence on the rent seeking behavior of NPEs and a natural policy lever to mitigate the impact of NPE activities. In fact, given the differing results on NPEs and practicing entities, policies that deal with vague patent language may be able to achieve their goals without disrupting the wider IP system.

In terms of policy, our quantitative estimates suggest that improving the patent examination process, in particular by ensuring consistency in the narrowing and clarification of claims across examiners, has potentially very large returns. Indeed, a lower bound based on the estimates presented in Sections 4 and 5 is that the share of NPE patents among granted patents could be reduced by 20% by implementing a "second pair of eyes" policy. As mentioned earlier, this calculation is based on litigation costs only, which are likely to underestimate the total social costs of NPEs' activities.

Future research could build on our methodology in various ways. Viewed at a high level, our methodology allows us to indirectly infer rent-seeking asset purchase behavior of certain agents in a system. This is done by exploiting random assignment of the co-producer of an asset, in our case the examiner. The conditions here are to our advantage in that we have two clear measures of co-producer tendencies that are orthogonal or negatively correlated with technological value, which we exploit to show that these co-producers are more likely to have their assets purchased by certain agents, in this case NPEs. We hope to see future work in this spirit, using asset production structure to test for rent-seeking behavior in other asset markets. Another possibly fruitful area of research is to test existing patent and innovation theory by leveraging our methodology to infer characteristics of patents. This could include looking at other ways patents are used in the IP system at large (such as licensing and commercialization of patent-protected products), and testing theories that predict the effects of patent breadth on real outcomes such as follow-on innovation.

4. The Lifecycle of Inventors[1]

4.1. Introduction

Innovation is at the center of growth theory and advanced economies use a broad range of policies intended to spur innovation, ranging from subsidies for research and development (R&D) to investments in technical education. However, relatively little is known about the characteristics and life trajectories of inventors. Indeed, information on even the most basic demographic characteristics - such as the age distribution and income composition of patent holders - is scarce because existing databases do not record such information.

In this paper, we present the first comprehensive portrait of inventors in the United States by linking data on all patents granted between 1996 and 2014 to federal income tax returns. Our linked dataset contains information on over 1.2 million patent applicants or holders. We use this linked data to document a set of stylized facts about the lives of "inventors"[2] that inform current theoretical debates and identify new patterns to be explained by the next generation of models of innovation and growth. We structure our analysis around the chronology of an inventor's life, starting with her family background and neighborhood at birth, then turning to her education and finally to her labor market career.

[1]Co-authored with Alex Bell, Raj Chetty, Neviana Petkova and John Van Reenen.

[2]Patents have well-known pros and cons as indicators of invention. Not all innovations are patented and not all patents correspond to meaningful innovations. But we use the rich information on the patent documents to deal with these drawbacks, for example, using future citations received by a patent as a proxy for its quality.

While our focus is primarily on descriptive facts rather than identification of causal mechanisms, the facts we document help discriminate between alternative theories in the literature and shed light on the types of policies that are likely to be most effective in sparking innovation. Most economic research on innovation policies focuses on what could be termed the "intensive margin" of innovation , namely getting more innovation out of the existing stock of inventors. For example, the US Research and Experimentation Tax credit reduces research costs relative to other forms of investment. Similarly, a leading argument for low top income tax rates is that they increase incentives for innovation and therefore boost growth (e.g. Mankiw (2013)). One potential problem with such policies is that they rely on those with the current potential to innovate to do more of it.[3] In light of these limitations, Romer (2000) recommends studying policies that focus on increasing the "extensive margin" of innovation - increasing the underlying supply of inventors. Our analysis of the lives of inventors yields a better understanding of what makes and constrains a potential innovator, which is an essential ingredient for developing such policies.

We begin by studying the birth and origin of potential inventors. The children of high-income parents are much more likely to be inventors: children born to parents in the top 1% of the income distribution are more than ten times as likely to become an inventor as children born to families with below-median income. Part of the relationship between parent income and children's patent rates could stem from children of the rich being born with higher ability than those of the poor and therefore being naturally more likely to make technological breakthroughs. Alternatively, part of the inventor-income relationship could be that even when children begin with identical traits, having low-income parents may hold children back from becoming innovators because of a lower quality education, neighborhoods, mentors, or jobs opportunities. If such barriers are important, this is very policy-relevant.

[3] In fact, there is a risk that with an inelastic supply of innovators, subsidies to research will simply increase the equilibrium wage of R&D scientists, rather than stimulate a greater volume of innovation (e.g. Goolsbee (1998)).

It would mean a lack of opportunity for those who would have had a comparative advantage in innovation but for their financial position. It would mean a loss of innovation and output due to a misallocation of talent. How many "lost Einsteins" could there be due to inequality of opportunity (e.g. Celik (2014))?

We shed light on these alternative mechanisms by using data from all individuals who went through the New York City (NYC) public school system between 1989 and 2009, from which we have standardized data on test score results in grades 3 through 8. We show that only around 30% of the invention gap between rich and poor can be accounted for by third grade math test scores. We conduct a similar analysis to study the gap in innovation by gender and by race. Unlike the gap in innovation by parents' socio-economic status, only 3% of the gender gap in innovation can be explained by differences in math test scores. Comparing Blacks and Hispanics to Whites, we also find a large invention gap - only a small fraction of it is due to initial ability, similar to the gender gap. A much more substantial fraction is due to income differences.

Existing talent misallocation models such as Hsieh et al. (2013) are based on Roy models of occupational choice according to comparative advantage but with frictions that create additional costs to all those from disadvantaged groups. These "rational sorting" models imply that individuals from disadvantaged groups who do become inventors should have *higher* levels of human capital than their more advantaged counterparts. In fact, our data shows the opposite - inventors from disadvantaged groups do *not* appear more talented. The quality of their patents (as measured by citations) and their initial test scores are similar or worse than other inventors.

What alternative model could account for these findings? We emphasize two related phenomena at play as a child grows up prior to choosing a career. First, using the NYC data on test scores from grades 3 to 8 we show that a substantial difference in educational outcomes between rich and poor families opens up as children progress through school. Using a wider sample we show that by the time we know the identity of the college attended, there is relatively little difference in whether a rich or poor child becomes an innovator. Second, we show that "exposure" to innovation in childhood has a strong association with the chances of growing up to be an inventor. Our measures of exposure include (i) whether the parent was an inventor; (ii) how innovative was the industry where a child's parents worked and (iii) neighborhood characteristics such as innovation in the childhood Commuting Zone (CZ). Being exposed to innovation when a child is growing up is strongly associated with later becoming an inventor. On all of these measures we show that it is not only the amount of innovation, but exposure to *type* of innovation that matters by using detailed technology class information. It is not simply that children who grow up in the Bay Area are more likely to be inventors (even when they live elsewhere) - they are more likely to specialize in the technologies that are relatively successful in the Bay Area (like computer software relative to medical devices). This evidence suggests that mentoring effects or exposure to careers in science and innovation at young ages may play a key role in children's later outcomes. Since children from low-income backgrounds are less likely to benefit from such exposure, this evidence reinforces the view that the innovation gap between the rich and the poor is driven by differences in environment and human capital accumulation, not intrinsic traits.

In the last part of the descriptive life-cycle analysis we look at the labor market and show that the returns to innovation appear highly skewed and uncertain, especially at the time of career choice. Returns often come later in life and are earned not just after the patent event, but during a broader period of several years leading up to patenting.

Motivated by these findings, we present a simple inventor lifecycle model that has barriers to human capital acquisition as rational sorting models of misallocation, but extends such models to allow for imperfect information over inventor careers. In particular, we argue that many individuals from disadvantaged groups may under-estimate the net benefits of an inventor career if they are not exposed to innovation during childhood. The model's predictions broadly match the findings of our dataset in a way that existing models cannot. Simple calculations suggest that the returns of supply-side policies that would realistically reduce the innovation rate gap between privileged and disadvantaged groups would be extremely large, with the potential of increasing the population of inventors by over 30%. We also present a quantitative analysis of the effect of changing the top marginal tax rates on inventors' incomes.[4] We show that the effect of top tax rates on the key individuals in the innovation process - the inventors themselves - is likely to be small due to the skewness and randomness of the payoffs. Our contribution is to calibrate this response using the skewness of the empirical earnings distribution of inventors. We show that even large cuts in top income tax rates on inventors will only induce a small change in the population of inventors.

We emphasize from a positive perspective that "extensive margin" innovation policies drawing talented individuals into the innovation sector may be very effective at increasing innovation, given that many talented low-income individuals currently do not make the choice of becoming inventors, in part due to the lack of exposure to this occupation as they grow up. The question of the optimal allocation of talent across sectors extends beyond the scope of this paper. Although we discuss how our results relate to common mechanisms in the "misallocation" literature, we do not draw normative conclusions about the observed distribution of talent across sectors (in particular, whether or not it is optimal to find so few talented low-income individuals among inventors). Rather, our results show from a positive perspective that exposure to innovation is an important driver of occupational choice or, in other words, that it has an important "allocation" effect. We view this finding as an important lesson for innovation policy, which is currently overwhelmingly focused on "intensive

[4]Note that we only consider increasing top tax rates *on inventors*, while holding the tax schedule fixed for other agents in the economy. Our evidence does not shed light on broader effects of changes in tax rates on innovation dynamics. Indeed, beyond inventors, many other agents are involved in the innovation process, for instance firms and financiers, for whom the returns to innovation may not be analogous to a random draw (e.g. because they hold large and diversified portfolios of innovations). Moreover, our analysis doesn't take general equilibrium effects into account.

margin" incentives targeting individuals who are already part of the innovation sector. Naturally, exposure effects may be a key driver of occupational choice in other contexts as well (for example for doctors, lawyers, financiers, etc.). In this paper, we focus on the decision of becoming an inventor for two reasons. First, inventors play a key role for growth and determining which policies have the potential to allocate more individuals to this occupation is therefore potentially very important for social welfare. Second, studying inventors has various methodological advantages: we can precisely characterize exposure effects by exploiting variation across detailed technology classes, and the discrete nature of patent applications allows us to conduct events studies to measure inventors' financial returns to innovation.

Our findings contribute to several vast literatures. The theoretical literature on the individual incentives to innovate are summarized by Scotchmer (2004). For example, there is extensive work on how different types of employment contracts will alter innovation incentives (e.g. Pakes & Nitzan (1983); Franco & Mitchell (2008)). Second, there is a growing literature on how misallocation can be a first order constraint on economic performance (e.g. Hsieh & Klenow (2009)). Specifically, Hsieh et al. (2013) argue that 15-20% of US GDP per worker growth 1960-2008 can be explained by the improved allocation of talent by race and gender. We find evidence that innovation (and growth) are held back because highly talented children from low income families are not becoming innovators as quickly as their richer, but less talented, peers. Although the link between inter-generational inequality and misallocation has been frequently discussed, it has not to our knowledge ever been examined in a systematic statistical manner. We bridge the gap between the endogenous growth literature (**?**) and the reallocation literature (Hsieh et al. (2013)) by showing that allocation of talent affects the rate of innovation and therefore long-run growth. In contrast, the existing reallocation literature has focused on higher productivity levels as misallocation is reduced. Third, there is a literature on academic scientists (e.g. Azoulay *et al.* (2010a), Azoulay *et al.* (2010b)) where biographies can be more easily built up. There is related work looking at sub-sets of patentors, especially star scientists in bio- and nano-technology (e.g. Zucker *et al.* (1998)). In parallel work using the same patent- tax data merge, Jaravel et al. (2015) show that teamwork is critical in the typical inventor's career, with large spillover effects from peers across the skill distribution.

Finally, there is a literature on looking at the characteristics of patentors (see Jung & Ejermo (2014) for a survey). The classic study is Schmookler (1957) who examined 87 US patentors and the most comprehensive recent work is the PatVal-EU data (e.g. Giuri *et al.* (2007)) which covers 9,107 inventors filing at the European Patent Office. An issue with these studies is that sample responses are low and possibly non-random. In response to this issue researchers have recently started matching patent data to near population employer-employee administrative datasets.Toivanen & Vaananen (2012) match administrative wage data to 1,800 Finnish inventors at the US Patent Office (as do Depalo & Di Addario (2015) on matched Italian administrative data). Using this Finnish data combined with a distance to college instrument, Toivanen & Vaananen (2015) argue that access to schools offering post-graduate engineering training have a causal impact on becoming an inventor. Jung & Ejermo (2014) use Swedish patents to examine the issue of gender and age differences for just under 20,000 inventors and Dorner *et al.* (2014) match the IEB employer-employee data (see Card *et al.* (2013)) to a cross section of German patents in 2002. The advantage of our data over these complementary papers is that (i) it is far larger than these other studies being at least one order of magnitude larger; (ii) we focus on the US as the country that is at the technology frontier in most industries and (iii) the US also has a relatively competitive labor market and so is less likely to depend on institutional idiosyncrasies. Finally, in terms of substantive questions, none of these earlier papers has systematically investigated the relationship between parental income and children's subsequent innovation. To our knowledge the only other paper to do this is the excellent (and complementary) paper by Aghion *et al.* (2015b) which looks at a similar set of issues extending the rich Finnish data (which also has IQ data on the male population).

The structure of the paper is as follows. The next section describes the data, Section III presents initial characteristics in early childhood; Section IV looks at schooling and exposure during later childhood and section V examines inventors in the labor market. Section VI discusses the model, its relationship to our stylized facts and its policy implications. Section VII concludes.

4.2. Data

4.2.1. Patent Data

We combine two sources of raw patent data. First, we use the several thousand weekly text and XML files of patent grant records hosted by Google. The files on this page contain the full text of about 5 million patents granted from 1976 to today, extracted from the USPTO's internal databases in weekly increments. We focus on the 1.7m patents that were granted between 1996 and 2014 to US residents. Second, we use data on 1.6m patent applications between 2001 and 2012 (Strumsky, 2014).[5]We use the names of all individuals denoted as inventors in the patent documents, not just those who are also assigned the intellectual property rights (i.e. the "self-assigned" holders of the patent rights). For example, if an individual is working for a firm, it is usually the company who will be the assignee rather than the employee who will still be named as the inventor. We define an individual as an inventor if he or she is named as such on the patent application or grant, have a US address and applied for the patent in the 1996-2012 period (to match the IRS data).[6]

[5]In 2001 the US moved into line with other patent offices and published patent applications 18 months after filing. Prior to this only successful applicants who were granted patents had their details published. For a fee, applicants can choose to have their filing kept secret and 15% of applicants choose to do so. The analysis in Graham & Hegde (2015) suggests that these (non-granted) applicants were of very low value. We show below the robustness of the results to considering only granted patents.

[6]So if a patent applied for in 2012 (or earlier) is granted by 2014, the individual is classified as an inventor.

4.2.2. IRS Data

From Treasury administrative tax files, we collect information on inventors' city/state, employer ID and adjusted gross income, as well as their current citizenship status and gender sourced from Social Security records. Most data are available starting in 1996 (and currently ends in 2012). Wages and employer ID are available only starting in 1999.

4.2.3. Matching

We match inventors to taxpayers using inventors' name, city, and state. Any inventor whose given address is outside of the United States is excluded from the matching process and dropped. We find equivalent information for taxpayers on 1040's, W2's, and other information return forms. The iterative stages of the match algorithm are described in more detail in Appendix A. We match approximately 88% of inventors of patents applied for in the last decade (the period in which information returns are most available) and slightly above 80% in the late 1990s.

We conduct various exercises to assess the quality of the match using additional data sources (e.g. data on inventor age from Jones (2010)).[7] We also explore selection on observables and find no strong selection effects (see Appendix). Because the taxpayer data is a source of linked observations of variants of a person's name and cities of residence lived in over time, the matching process provides a simple way to link different patents filed by the same inventor even if the inventor's name differs across patents or he has moved cities. [8]

[7] Jones (2010) determines the ages of 55,000 inventors using name, zip code information, and a public Web site (http://www.AnyBirthday.com). Comparing the birth dates obtained from this website and from Treasury tax files for the inventors that are in both Jones' database and ours, we find close to an exact match.

[8] While the panel nature of the linked dataset allows us to see full income and patenting profiles for our sample, a drawback is that we cannot classify individuals as inventors who were active only prior to 1996.

4.2.4. Inter-Generational Analysis

For the analysis in which we study parental income, we can only look at a sub-sample of the IRS data for adults who are born after 1980, the earliest cohort for which we have sufficient records to match them to parents through 1040 forms that record dependents. For years when parents file tax returns, we calculate parents' income as the pre-tax household income; we gather parent income from W2 and other information returns in years when a parent does not file. Further details on the process of matching children to parents are outlined in Chetty *et al.* (2014b). Looking at the sample of individuals born 1980-84 who would be aged 28-32 in 2012 we still have a substantial sample of 45,083 inventors. This focus on "young" inventors may seem a disadvantage, but 13% of patents in our data are invented by those aged 32 or under. We also show our results are robust to using older cohorts from the Statistics of Income, which is a 0.1% sample of the IRS data available for years prior to 1996.

4.2.5. Test Score Analysis

When looking at the test scores of elementary school children, we further condition on a sub-sample in which we know whether children attended a New York City public elementary school between 1989 and 2009. We observe standardized test scores in grades 3 through 8 for these children, and limit our analysis to the approximately 250,000 of these students born between 1979 and 1985. Further details of the process by which these students were matched to taxpayers can be found in Chetty *et al.* (2014a).

4.2.6. Initial Data Description

Table A1 contains descriptive statistics. In the whole sample there are a total of 1.2m inventors in our matched data (Panel A). The innovation data is highly skewed - the average inventor has 3.3 patents with 28.5 citations. But the standard deviation is enormous: 9.1 for patents and 154.6 for citations. The annual wage of an inventor is just over $112,000 at the mean and $81,000 at the median, with incomes of over $178,000 (mean) and $109,000 (median). These are higher than for US workers as a whole. Consistent with Hunt (2009), only 11.6% are women. The average age of an inventor is 45.[9] Comparable data for the inter-generational analysis sub-sample and test score sub-sample are in Table A1 Panels B and C respectively.[10]

4.3. Birth and Early Experience

4.3.1. Parental income

Figure 1 shows (solid blue circles) the number of inventors per 10,000 individuals (left hand vertical axis) ordered by the parents' percentile position in the national income distribution (x-axis).[11] We measure the latter as average household income 1996-2000, which Chetty *et al.* (2014b) have shown to be a good proxy for permanent income. A sharp, quite convex upward slope is apparent. Children born to the richest 1% of parents had invention rates of 8.3 in every 10,000, which is an order of magnitude higher than the proportion of inventors born in the bottom half of the income distribution (0.85). One hypothesis is that kids from richer parents are more likely to produce more low-value patents, but the distribution of more

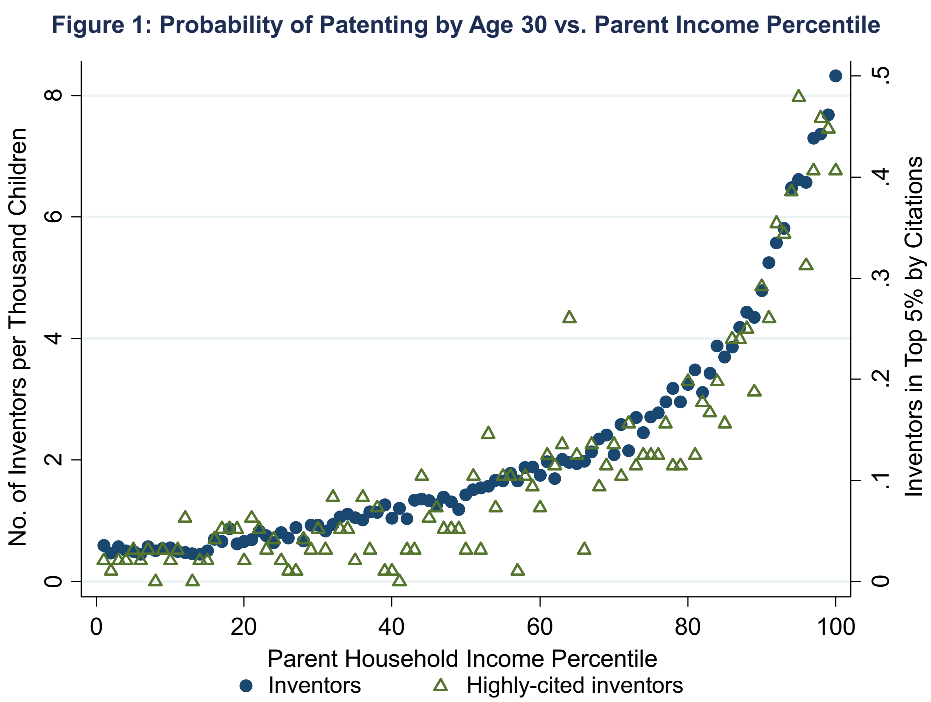
[9] 1.

[10] The low levels of income and wages are because these are younger people who will usually be earning no income while at school.

[11] We use the 1980-82 birth cohort, but similar results are apparent using the 1980-84 or cohorts from individual years.

prolific inventors is not dependent on parent income. To test this hypothesis, we repeated the analysis using as an outcome whether an inventor was in the top 5% of her age group's lifetime citation count. The green triangles corresponding to the right-hand vertical axis show that the positive relationship between parental income and invention is just as strong for high-quality patents as it was for all patents.[12] As noted above, we test for whether the results are specific to looking at young inventors by using the Statistics of Income 0.1% IRS sample. We take a cohort born in 1970-72 (ten years younger than those in Figure 1) and examine the fraction of inventors aged 30-40 (instead of aged 30-32). Figure A1 show that the strong gradient between parental income and inventor status is clearly visible in this older sample (although it is noisier due to smaller sample size).

Figure 4.1.: Probability of Patenting by Age 30 vs Parent Income Percentile



This inventor-parent income relationship has never been comprehensively documented before

[12]Similarly, Figure A2 shows that if we use alternative definitions of inventor status such as just patent grantees or just the post 2001 applicants data we again obtain a ratio of top 1% inventor rates to bottom 50% rates of about 10 to 1.

to our knowledge. But one could regard the relationship in Figure 1 as deeply unsurprising. We would expect that parental income was positively associated with many other indicators of "elite" success such as becoming a successful lawyer, physician, hedge fund manager or economist, for example. 9.7% of children born to the richest 1% of parents stayed in the top 1%, compared to only 0.3% of children in the bottom 50% getting into the top 1% (see Figure A3). We focus on inventors because there is ample evidence that there are positive spillovers from innovation and therefore the social returns are greater than the private returns.[13] Hence, understanding barriers to the creation of more inventors is more important for public policy than the supply of hedge fund managers or lawyers. Furthermore, the focus on inventors also enables us to implement empirical strategies, such as the use of detailed technology class information in patent documents, to shed light on whether the relationship in Figure 1 is related to the childhood experience of potential innovators. Such strategies would be harder to implement with other professions.

There could be many reasons for the relationship in Figure 1. For example, since income is related to ability and this human capital is partly genetic, the relationship could be due to inherited ability. Alternatively, it may be that a poorer child begins life equally gifted as a richer one, but faces barriers to becoming an inventor. To investigate these issues we turn to the New York City (NYC) schools data where we know standardized statewide test scores of children in grades 3 to 8 in math and English. Figure A4 shows that the inventor-income gradient holds in this sub-sample (cohorts born in 1979-84). We use a parental income split at the 80th percentile and label the "rich" those above this and the "poor" those below. This is somewhat arbitrary, but other parental income splits produce similar qualitative results to everything we will show below.[14]

Figure 2 shows the kernel density of third grade math test scores for rich and poor kids.[15]

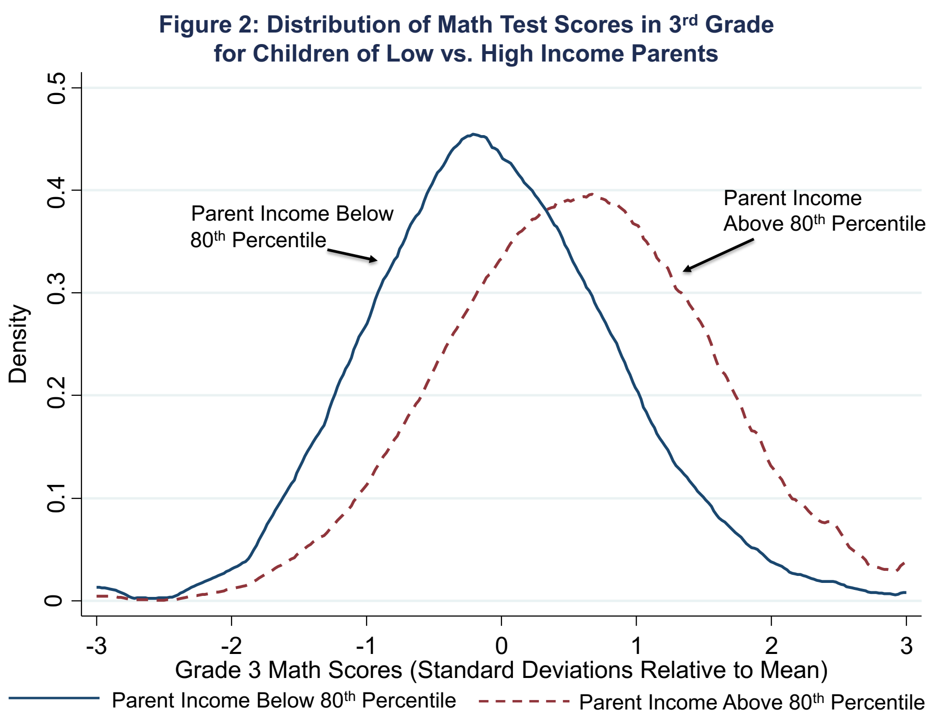
[13]For example, Bloom *et al.* (2013),Jones & Williams (1999) andGriliches (1992))

[14]All results available upon request.

[15]We are certainly not claiming that all third grade math scores are genetic. Rather we are using this as

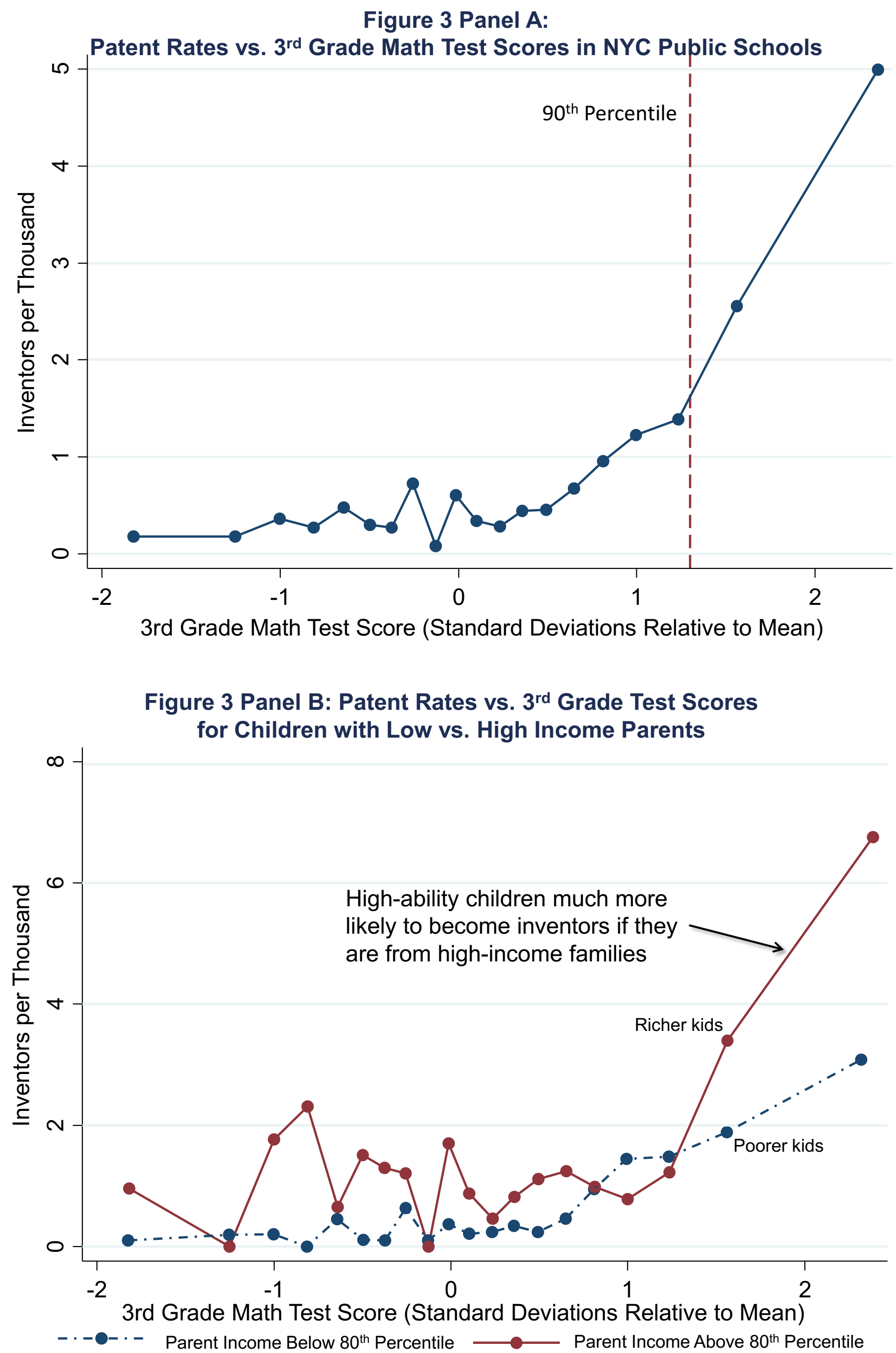
The rich kids' distribution is strongly shifted to the right as we would expect. Only 7% of poor kids are in the top decile of the math score distribution compared to 23% of rich kids. Panel A of Figure 3 shows the proportion of children who grow up to be inventors as a function of their math test scores in third grade. The pattern is striking. Children in the top 5% of the math distribution have a future innovation rate of over 5 in a 1,000 people whereas those in the next 5% have innovation rates of just over 2.5. Those in the bottom 90% have innovation rates of around 0.5.

Figure 4.2.: Distribution of Math Test Scores in 3rd Grade for Children of Low vs. High Income Parents



an indicator of early childhood environment and initial ability, which we want to distinguish from later childhood environment as children grow up.

Figure 4.3.: Probability of Patenting by Age 30 vs 3rd Grade Math Test Scores



These results can be shown in a regression setting (Table A2) where the dependent variable is whether the child grows up to be an inventor (scaled by 1,000). Column (1) shows that third grade math scores are highly predictive of becoming an inventor - a standard deviation increase in math test score is associated with a (highly significant) increase of 0.85 in 1,000 chance of becoming an inventor. Column (2) substitutes standardized grade 3 English test scores which enters with only a slightly lower coefficient of 0.68. Column (3) includes both math and English together and shows, interestingly, that conditional on math test scores (which remain highly significant) English test scores are insignificant. Column (4) has a less parametric version of column (3) which uses dummies for each vingtile of the English score and vice versa in column (5). Whereas math scores remain positive and significant in column (4) English scores are completely insignificant in column (6). This implies that early math ability is very informative for future inventor status whereas English performance is not. Note that this is not the case when re-estimating these equations but using a dummy for whether a child ended up in the top 1% of the income distribution as an outcome. The final three columns use this as the dependent variable and show that both English and math are significant in such a regression.[16]

Panel B of Figure 3 shows that there is a positive relationship between early math test scores and the chances of becoming an inventor for both rich and poor children and again, the relationship is noisy until we get to the top decile of the math test score distribution. It is striking, however, that for children in this top decile, rich kids have a much higher invention rate than poor kids. About 7 in every 1000 rich kids in the top 5% of math test scores become inventors whereas the rate is less than half this for poor kids. This strongly suggests that the innovation-income gradient cannot solely be attributed to early test scores.

We can quantify the role of early test scores in accounting for the innovation-income relationship in different ways. Table 1 (Panel A) provides a calculation based on the full distribution

[16] These results are also robust if we instrument one test score with another to address the issue of noise in the test scores. See Kahneman & Ghiselli (1962) for a discussion.

decomposition methodology of DiNardo *et al.* (1996) ("DFL"). We give the children below the 80th percentile of parent income the same math test scores as those from richer families (specifically, we divide the distributions into 5% vingtiles bins to do this). It is important to do this across the whole distribution rather than just at the mean (as in a conventional Oaxaca-Blinder decomposition) because most of the increase in innovation probability is in the right tail of the ability distribution as already shown in Figure 3. We calculate that the test score difference in third grade accounts for just over 30% of the difference in innovation propensities in later life. So this is a sizable proportion, but obviously there are other factors that must account for the majority of the difference. [17]

Table 4.1.: Fraction of Gap in Patenting by Parental Income Accounted for by Differences in Math Test Scores

Table 1: Fraction of Gap in Patenting by Parental Income accounted for by differences in Math Test Scores

Panel A: Decomposition Using 3rd Grade Math Score		
	Patent Rate (per 1000 Individuals)	**Gap Relative to Above p80 Group**
Above 80th Percentile of parent income	1.93	
Below 80th Percentile	0.52	1.41
Below 80th Percentile.(Reweighting Scores)	0.95	0.97
% of gap accounted for by 3rd grade scores		30.9% (s.e. = 8.5%)
Panel B: Decomposition using later grades		
% of gap accounted for by 4th grade scores		36%
% of gap accounted for by 5th grade scores		39%
% of gap accounted for by 6th grade scores		45%
% of gap accounted for by 7th grade scores		51%
% of gap accounted for by 8th grade scores		53%
Average percentage point change per grade		4.6

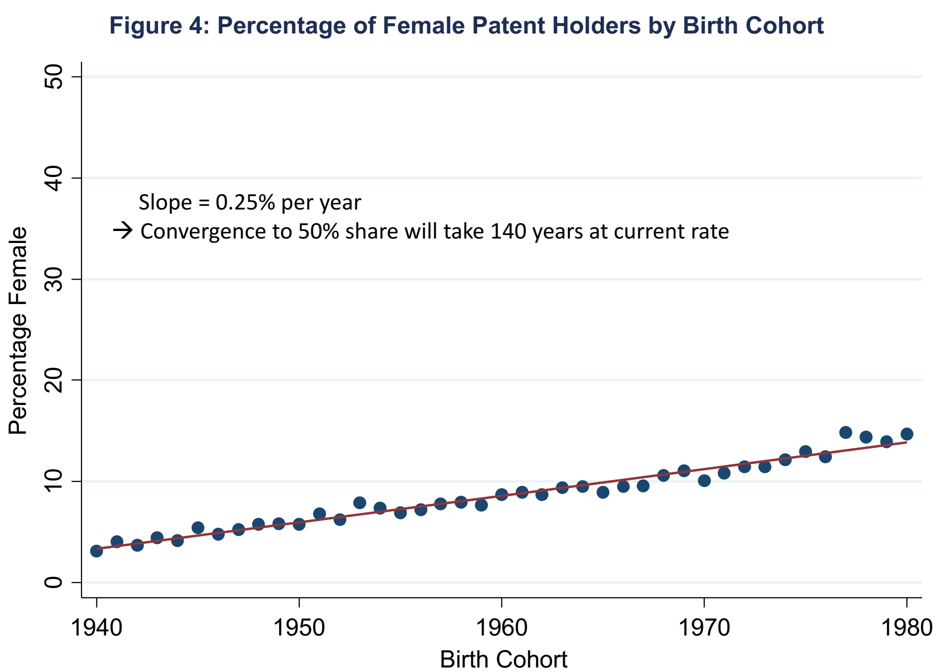
Notes: This is a DiNardo, Fortin and Lemieux (1996) decomposition of the difference in innovation rates among children of the "rich" (top quintile of income) and all others (see text) using the NYC test score data combined with our IRS-USPTO match. We divide the math distribution into vingtiles and give the low-income Children the test score distribution of the rich and re-calculate the implied innovation rates (re-weighting scores).

[17] In Table A3 we show several alternative ways of quantifying the contribution of test scores to the inventor-income gradient which reach broadly similar conclusions. Looking across the first row for third grade test scores using the balanced panel leads to a contribution of 35.5% (a bit higher than our baseline contribution of 30.1%). Using a split at median parental income (instead of the 80th percentile) leads to a smaller, 27.8% contribution. The last three columns use a method of introducing 20 dummy variables for each vingtile of the test score distribution and observing by how much the coefficient falls on the "rich" income dummy variable. This produces somewhat larger estimates of 49.4% in the baseline, 41.5% for the balanced panel and 43% for splitting at median income.

4.3.2. Gender

Our data shows that only a small proportion of inventors are female. In Figure 4 we show the fraction of female inventors by birth cohort. Clearly more women are becoming inventors over time: only 3% of inventors born in 1940 were female, whereas this had risen to 15% for the 1980s birth cohort. This is a very slow rate of convergence, however. Extrapolating this line forward suggests it will take 140 years before women reach parity with their male counterparts. There has been much recent discussion about the causes of this difference and its persistence.[18]

Figure 4.4.: Percentage of Female Patent Holders by Birth Cohort



[18] See Hunt (2009), Thursby & Thursby (2005), or Ding *et al.* (2006). There has been much controversy over this reported in the media. For example, the Nobel Laureate Tim Hunt argued in 2015 that women did not work well in the high-pressure culture of academic R&D labs (http://www.dailymail.co.uk/news/article-3117648/Ban-women-male-labs-distracting-cry-criticised-says-Nobel-prize-winner-Sir-Tim-Hunt.html). Larry Summers speculated that the lower proportion of female elite scientists could be because there was a greater variance in men's intrinsic ability than in women's (e.g. http://blogs.scientificamerican.com/the-curious-wavefunction/why-prejudice-alone-doesnt-explain-the-gender-gap-in-science/).

Table 4.2.: Fraction of Gap in Patenting by Gender Accounted for by Differences in Math Test Scores

Table 2: Fraction of Gap in Patenting by Gender accounted for by differences in Math Test Scores

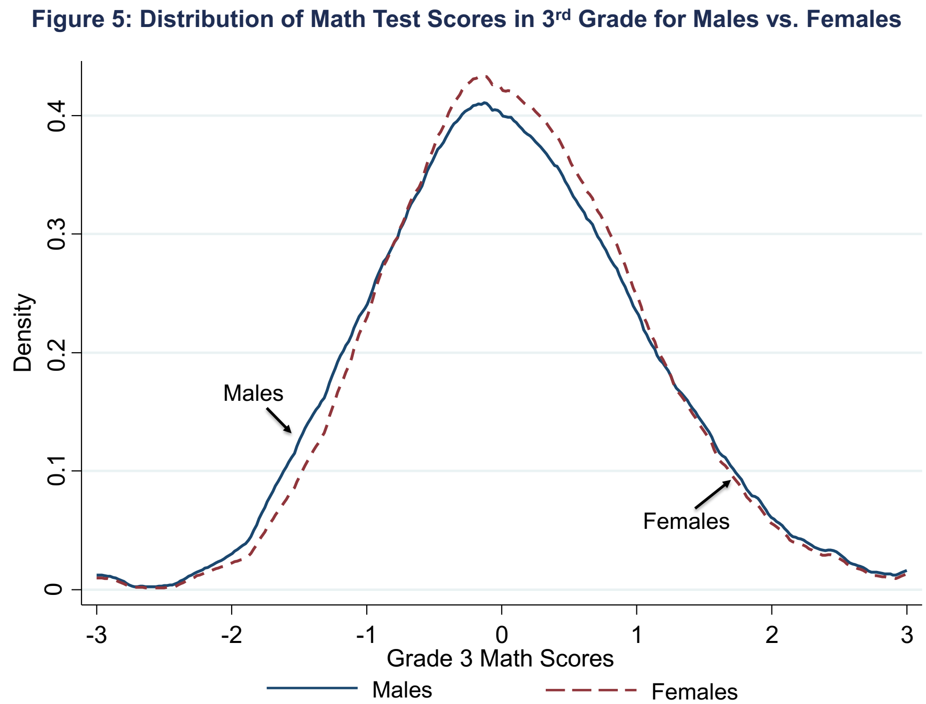
Panel A: Decomposition Using 3rd Grade Math Score	**Patent Rate (per 1000 Individuals)**	**Gap Relative to Above p80 Group**
Men	1.36	
Women	0.47	0.89
Women (Reweighting Scores)	0.50	0.86
% of gap accounted for by 3rd grade scores		3.7%
Panel B: Decomposition using later grades		
% of gap accounted for by 4th grade scores		2%
% of gap accounted for by 5th grade scores		2.9%
% of gap accounted for by 6th grade scores		4.5%
% of gap accounted for by 7th grade scores		6.4%
% of gap accounted for by 8th grade scores		8.7%
Average percentage point change per grade		1.7%

Notes: This is a DiNardo, Fortin and Lemieux (1996) decomposition of the difference in innovation rates among boys and girls (see text) using the NYC test score data combined with our IRS-USPTO match. We divide the math distribution into vingtiles and give the girls the test score distribution of the boys and re-calculate the implied innovation rates (re-weighting scores).

Figure 5 illustrates the relative similarity of the distribution of math test scores across both genders. Boys have a slightly higher score at the mean and are slightly thicker in both tails. [19] We conducted an analogous decomposition exercise for women as we did for parental income in Table 2. This shows that we can account for very little of the gender-innovation gap using math test scores. We account for nothing at all using third grade test scores and only about 3% using eighth grade scores. So, in stark contrast to income, the gender gap in innovation does not appear to be ability related.

[19] This is consistent with Machin & Pekkarinen (2008), who show that boys' test scores had significantly higher variance than girls' scores. They examine the OECD's standardized PISA math and reading tests taken by 15 year olds. In 37 of the 41 countries examined the boy-girl variance was significantly different for math.

Figure 4.5.: Distribution of Math Test Scores in 3rd Grade for Males vs. Females



4.3.3. Race

There is a small literature documenting racial differences in invention rates by race (e.g. Cook & Kongcharoen (2010)). Figure 6 uses the NYC data (where we can observe race) to show that there are wide disparities in patenting races by minority status.[20] The first blue bar shows that white children have an inventor rate of 1.6 in 1,000, which is more than three times the rate for black kids (0.5) and eight times the rate for Hispanics (0.2). By contrast, Asian children are twice as likely to grow up to be inventors than whites (an inventor rate of 3.3). We can implement the DFL decompositions to see how much of these differences can be accounted for by third grade math test scores. The second red bar in Figure 6 does this for each racial group where we take white kids as the base and normalized to 1.6. We can see that the bars are not changed very much by this reweighting, with each gap shrinking by

[20] The share of students by race is: Asian 7.46%, Hispanic 32.81%, black 38.95%, white 20.78%.

only 0.1. For example, the Black-White gap shrinks from 1.1 to 1.0, a change of under 10%. By contrast, re-weighting by income (the third green bar) makes a much bigger difference with the Black-White gap falling by almost half from 1.1 to 0.6. This suggests income is much more important than ability in accounting for the inventor difference between blacks and whites. Income makes less difference for the white-Hispanic gap, however, falling from 1.4 to 1.3. The White-Asian gap actually widens from 1.5 to 2.6 when we reweight by income as Asian parents in NYC public schools are on average poorer than white parents. Figure 7 shows the race results in another way. If we plot the inventor-test score gradient by racial group there are hardly any differences except in the top 15% of the ability range. It is amongst the most gifted at math that the differential invention rates by race become clearly visible. For third graders in the 10% of the math test score distribution, future inventor rates are over 8 for Asians, about 4 for Whites and about 1 for Blacks and Hispanics.

Figure 4.6.: Patent Rates by Race and Ethnicity in New York City Public Schools

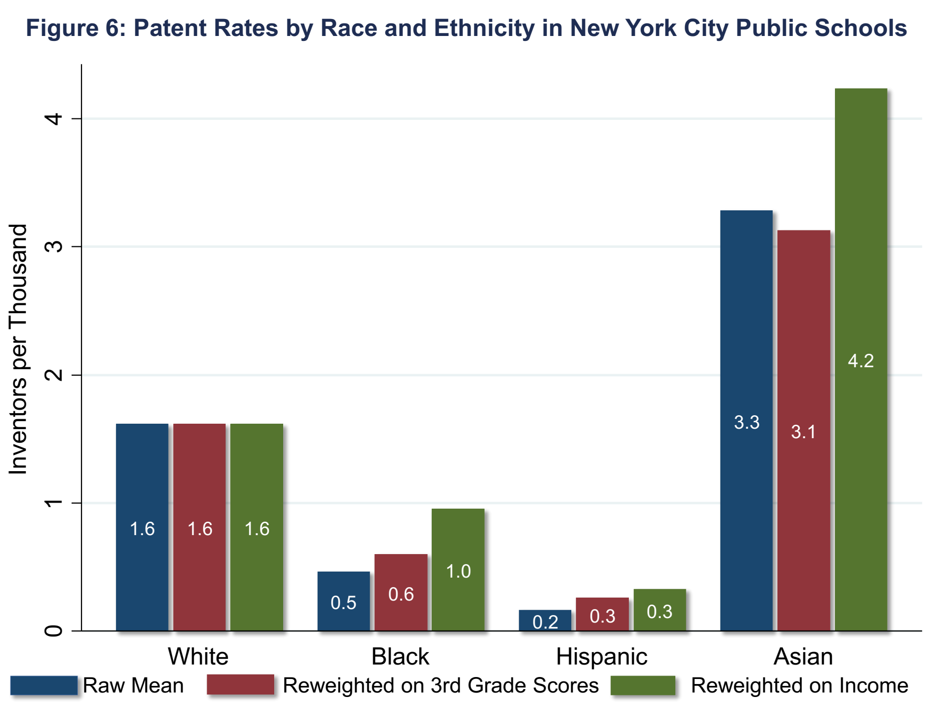
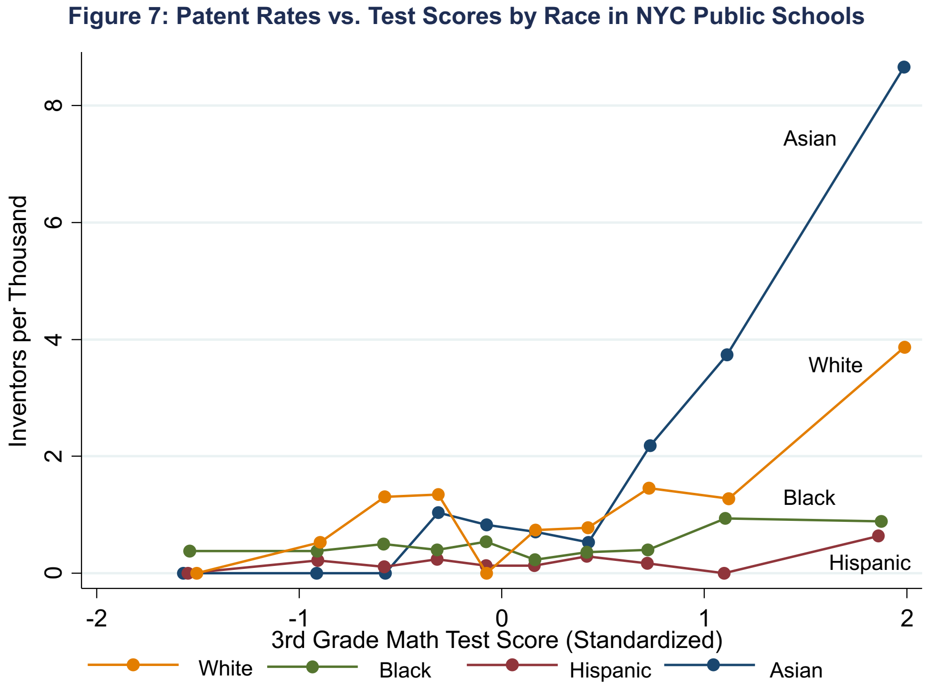


Figure 4.7.: Patent Rates vs. Test Scores by Race



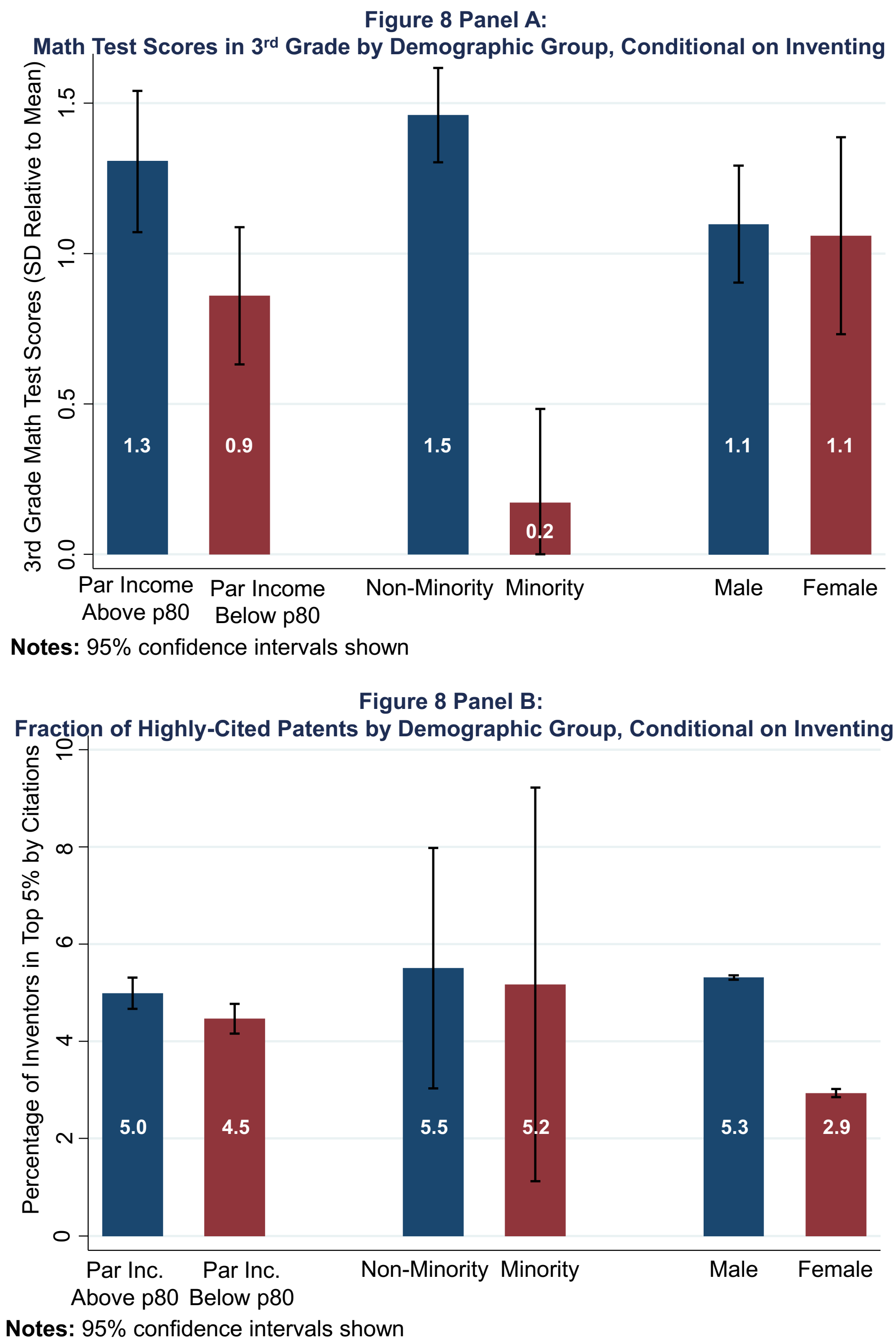
4.3.4. Implications for Models of Talent Allocation

One interpretation of these differences in invention rates by income, race and gender is that they reflect barriers to becoming an inventor over and above the intrinsic ability and preferences of individuals. In a recent contribution, Hsieh et al. (2013) develop a Roy model of heterogeneous occupations which have barriers to entry that differ by group (e.g. race and gender). These barriers or frictions are a combination of direct discrimination, in which individuals are paid less than their marginal product, and barriers to the acquisition of the type of skills that are useful to enter the occupation (e.g. a legal training to become a lawyer). This causes a misallocation in which a talented individual may not sort to the occupation that best fits her comparative advantage. The frictions cause misallocation and a loss of welfare. Calibrations in Hsieh et al. (2013) suggest that improvements in the allocation of talent (primarily from less barriers to women) were responsible for 15% to 20% of the growth in US output per worker between 1960 and 2008.

One might, however, be skeptical over whether these type of rational sorting models will generate first-order welfare losses. Consider such a model applied to our context of inventor careers. Suppose that children from poorer families have difficulties in accessing high-quality schools to improve their math ability, increasing the cost for them to become successful inventors. To the extent that such barriers exist, rational sorting models imply that they will dissuade only marginal inventors. The most talented individuals will still be prepared to make the effort to acquire skills necessary to become an inventor. It is not "Einsteins" who are lost to the R&D sector, but rather those of more mediocre ability. By contrast, we discuss below a model where disadvantaged groups are relatively less informed and underestimate the net benefits of an inventor career. In this case, even some very talented potential inventors will not pursue a career in innovation because they miscalculate the true expected cost-benefit ratio. There could be more first-order welfare losses in such a model as the breakthroughs of possible Einsteins may be lost.

To investigate whether the rational sorting model is the most plausible explanation of the patterns we observe, we implement a simple non-parametric test. This class of models has a clear prediction that conditional on becoming an inventor, individuals from the discriminated group should be, on average, of *higher* talent than the non-discriminated group. We investigate this in two ways in Figure 8. First, we use our third grade math test scores as a measure of ability. In Panel A we show the mean math grades for inventors split into (a) whether their parents were rich or poor, (b) whether they were white or from a minority group (black or Hispanic) and (c) whether they were male or female. The results are striking - in no case are the mean test scores higher for the "discriminated group" as rational sorting models would suggest. If anything, it is the opposite with rich kids and whites having significantly higher scores conditional on being an inventor (the male-female gap is insignificant).

Figure 4.8.: Math Test Scores Conditional on Inventing



The second empirical way we implement our test of rational sorting is to use patent citations as a measure of inventor ability.[21] We define a highly cited patent as one which is in the top 5% of future citations for its cohort (as in Figure 1). Again, we find no evidence for the basic rational sorting model, which predicts that the quality of innovation should be higher for the disadvantaged group. Conditional on patenting, mean citation rates are similar (or higher) for men, whites and those from high-income backgrounds.[22]

The upshot of this discussion is that the rational sorting models may be underestimating the loss from misallocation in our context. To consider some alternative explanations of how the allocation of individuals across sectors can occur, we next turn to the conditions under which inventors grew up.

4.4. Inventors Pre- Labor Market

We look at two aspects of the period of time between early school and joining the labor market for potential inventors: direct measures of later schooling (sub-section IV.A) and then "exposure" to innovation (sub-section IV.B).

4.4.1. Schooling

To investigate the importance of human capital acquisition we repeated the DFL decompositions in Table 1 but exploit information from later grades (NYC data goes through grade 8). Panel B shows the DFL results. As children get older, test scores account for more of the inventor-income gap. By 8th grade 53% of the gap is accounted for, compared to 30%

[21] A disadvantage of this first test is that it might be privately optimal for high-ability disadvantaged groups to go into other high-skilled occupations such as investment banking (we do not have an early measure of inventor-specific ability). The second test does not suffer from this drawback.

[22] Note the smaller confidence bands on gender is because in Panel B we can use the entire matched IRS-patents database whereas in Panel A we just use the NYC data as we are restricted to using test scores. The large confidence bands for minorities is because there are very few highly cited patents in these cells and because minority status is only available for the NYC sample.

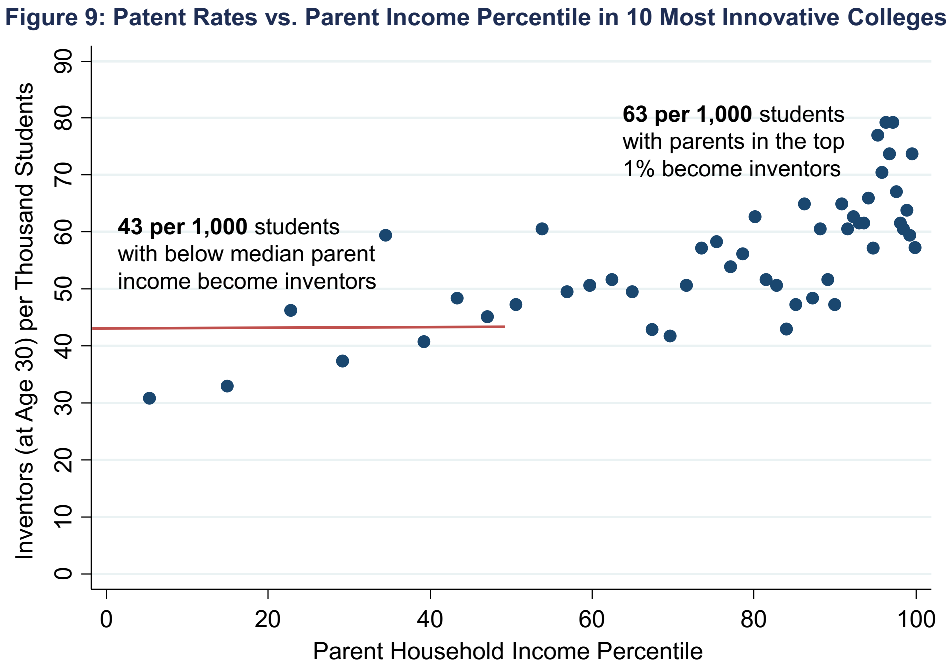
in grade 3. On average an extra 4.4 percentage points of the gap is accounted for each year by test scores and the null hypothesis that there is no additional explanatory power of the later grades is rejected (p-value = 0.025).[23] Low-income kids fall further behind richer ones as they progress through school. In in Panel B of Table 2 we present decompositions for gender. Although test scores do become more important as children pass through school, the additional explanatory power of schooling is very low in explaining the gender gap for innovation. By 8th grade, test scores still only account for under 9% of the advantage of boys over girls. This suggests that explanations for gender differences in invention are not ability-related, at least as measured by test scores.

For the individuals who attended a US college between 1999 and 2012 (200,000 inventors) we can determine which college they attended. One striking fact is that a few select colleges produce an enormous fraction of US inventors. For example, the top 10 most innovative colleges (defined as the colleges whose graduates will be granted the most patents) account for only 3.7% of enrollments, but 15% of (citation-weighted) patents (Figure A5). In these top 10 colleges, 5.6% of students have a patent by age 30 (about 28 times the average for the population as a whole by this age). In the 10 most innovative colleges, the invention-parental income relationship is severely attenuated. Figure 9 shows that in this sample children whose parents were in the (nation-wide) top 1% of the income distribution produced 63 patents per 1,000 (compared to 0.1 per 1,000 in Figure 1), compared to 43 per 1,000 in the bottom half of the income distribution. This ratio of about 1.5 to 1 compares to 10 to 1 in Figure 1. Hence, by the time students graduate from a school like MIT or Stanford, the income of the student's parents makes relatively little difference. The main role of higher income is increasing the chance of a child going to one of these colleges - Table A7 shows the sharp positive gradient between being born into a wealthy family and attending an innovative

[23]The lower rows of Table A3 show this is equally true using other decomposition methods (up to 61% by grade 8 in the median split specification of column (6)).

college.

Figure 4.9.: Probability of Patenting by Age 30 vs Parent Income Percentile in 10 Most Innovative Colleges



4.4.2. Exposure to Innovation

We next explore the role of being *exposed* to innovation in childhood and growing up to be an inventor. We examine exposure along three dimensions: (a) whether an individual's parent was an inventor (sub-section IV.B.1), (b) for children of non-inventors, what was the innovation in the industry of one's father (sub-section IV.B.2) and (c) innovation in the area where the child grew up (sub-section IV.B.3). For all three dimensions we examine not just the overall level of innovation, but the *type* of innovation by exploiting the information on patent technology class. Patents can be classified into seven broad categories,[24] thirty

[24] The classes are Chemicals; Computers; Communications; Drugs and Medical; Electrical and Electronic; Mechanical and "Others".

seven sub-categories and four hundred and forty five classes. Looking at whether exposure to specific technological fields (rather the innovation in general) is correlated with becoming an inventor in that specific field is a sharper test of whether becoming an inventor is affected by the environment rather than just being an innate trait.

4.4.2.1. Parental Innovation Status

We analyze innovation rates for 16 million children born between 1980 and 1984 for whom we know whether their parents themselves filed a patent since 1996. Amongst those whose parents were inventors, the patent rate was 11.1 per 1,000. By contrast, if a child's parent was not an inventor then the patent rate was only 1.2 per 1,000. A fraction of these were children-parent teams on the same patent. However, even if we remove these observations, then it is still the case that the inventor rate is 8.5 for the children of inventors.

Of course, this relationship within the family could reflect a genetic predisposition to be an inventor. To address this we examine the patent class in which the parent invented: it is very unlikely that there is a gene that codes specifically for a type of technology class such as "modulators" (technology class 332) vs. "demodulators" (technology class 329) or synthetic resins (class 520) vs. natural resins (class 530). Conditional on inventing any technology, children of inventors are nine times more likely to invent in the same sub-class as their inventing fathers than they would be by random chance.[25]

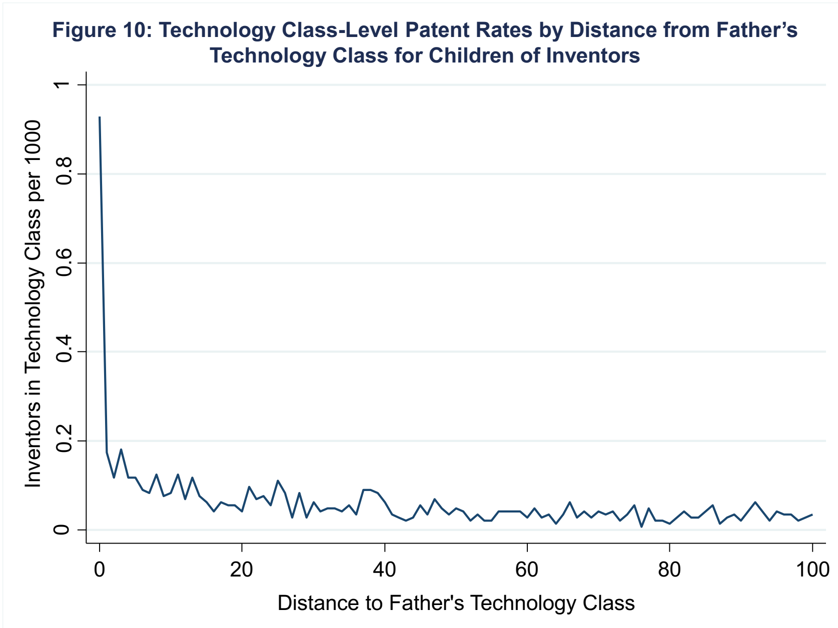
We also develop a closeness measure based on looking at whether a patenter in class A also patented in class B. This exploits the fact we have individual identifiers due to the IRS match and can examine cross-class patenting by individuals over time.[26] Table A4 gives an example. Starting with technology class 375 ("pulse or digital communications") it has

[25] 0.9% of children invent in the same class as their fathers compared to an expected 0.1% by random chance (see Table A9).

[26] Other distance metrics in the literature include classes within the same sub-category and cross class citations (see Bloom *et al.* (2013).

a distance of zero with itself by definition. For inventors who had a patent in pulse or digital communications, the next most popular class to patent in was demodulators so we give this a distance of 1. After this comes modulators (distance = 2) and so on. Using this information, Figure 10 shows the proportion of children of inventors patenting in the same technology class and then in other "close-by" classes. We calculate the proportion for each technology class and then average across all technology classes with weights proportional to the number of patents in each technology class. There is a clear spike in the probability of children of inventors inventing in the same technology class as their parent. On average, if an individual's parent invented in a particular technology class, the child had over a 0.9 in a 1,000 chance of inventing in exactly the same class. The chance of inventing in the next "closest" technology class (at distance = 1) was under 0.2, about a fifth as high. As the class becomes more distant, the child's probability of investing in that class diminishes (graphically, the downward slope going from left to right of Figure 10). This evidence is consistent with the view that being brought up in a family where there is some very specific knowledge about a technology helps the younger member of that family to go and become inventors in that technology.

Figure 4.10.: Technology Class-Level Patent Rates by Distance from Father's Technology Class for Children of Inventors



4.4.2.2. "Mentors": Innovation in father's industry

The direct relationship between parent's and children's inventor class is striking, but of course this is focused on a narrow slice of the data: those with parents who have invented. Our second measure of exposure looks at characteristics in the (six-digit) industry where a child's father worked. The idea is that the network of people in the firm and industry could influence what careers young people are interested in studying and pursuing in later life. Table 3 puts this results in a regression framework, where the data is constructed solely from all children whose parents were *not* inventors (in order to rule out the direct channels examined in the previous sub-section). Column (1) is run at the six-digit industry level. For each individual we calculate the proportion of workers in their father's industry who were inventors (right hand side variable). The dependent variable is the proportion of children who became inventors (within a father's industry). There is a strong positive and significant relationship between these two variables.[27] A one standard deviation increase in the fraction of inventors in the father's industry (0.0023) is associated with an increase of 0.0006 inventors, or 25.3% at the mean of the dependent variable (0.0023).

Table 4.3.: Children's Patent Rates vs. Patent Rates in Father's Industry

Table 3: Children's Patent Rates vs. Patent Rates in Father's Industry

Dependent variable:	(1) Fraction Inventors	(2) Fraction Inventing in Category	(3) Fraction Inventing in Sub-Category	(4) Fraction Inventing in Class	(5) Fraction Inventing in Class
Fraction Inventors in Father's Industry	0.250*** (0.0276)				
Fraction in Category in Father's Industry		0.162*** (0.0166)			
Fraction in Sub-Category in Father's Industry			0.154*** (0.0168)		
Fraction in Class in Father's Industry				0.0780*** (0.0136)	0.0601*** (0.0129)
Fraction in same Sub-Category but other Class					0.00438*** (0.00083)
Fraction in same Category but other Sub-Cat.					0.00006 (0.00040)
Fraction in other Category					0.00021*** (0.00005)
Cells	Father's industry	Father's industry*category	Father's industry*sub-category	Father's industry*technology class	Father's industry*technology class
Observations	345	2,415	12,765	153,525	153,525

Notes: Standard errors are clustered by 345 industries. Column (2) includes 7 category fixed effects; column (3) includes 37 sub-category fixed effects; columns (4) and (5) include 450 technology class fixed effects. The sample is children whose parents are not inventors.

In column (2) of Table 3 the unit of observation is an industry (345) by patent category (7) cell. We construct the data in the same way as before but calculate the proportion of inventors within an industry-category cell and then include category fixed effects in the regressions. Column (3) goes one step further and constructs cells defined by industry and sub-category (including 37 sub-category fixed effects in the regression) and column (4) is the most disaggregated using industry by technology class cells (with 450 technology class fixed effects). The same pattern emerges in columns (2)-(4) as in column (1). Children are much

[27] Figure A10 has a the scatterplot of this relationship.

more likely to innovate within a narrowly defined technology area if their father has worked in an industry that has more inventors in this same area. Column (5) replicates column (4) but includes (i) the fraction of inventors in the same sub-category but in a different industry class and (ii) the fraction in the same sub-category but a different class and (iii) other categories. All coefficients in these variables are positive and two are significant (the own class coefficient drops slightly but remains significant).[28]

4.4.2.3. Innovation in childhood neighborhoods

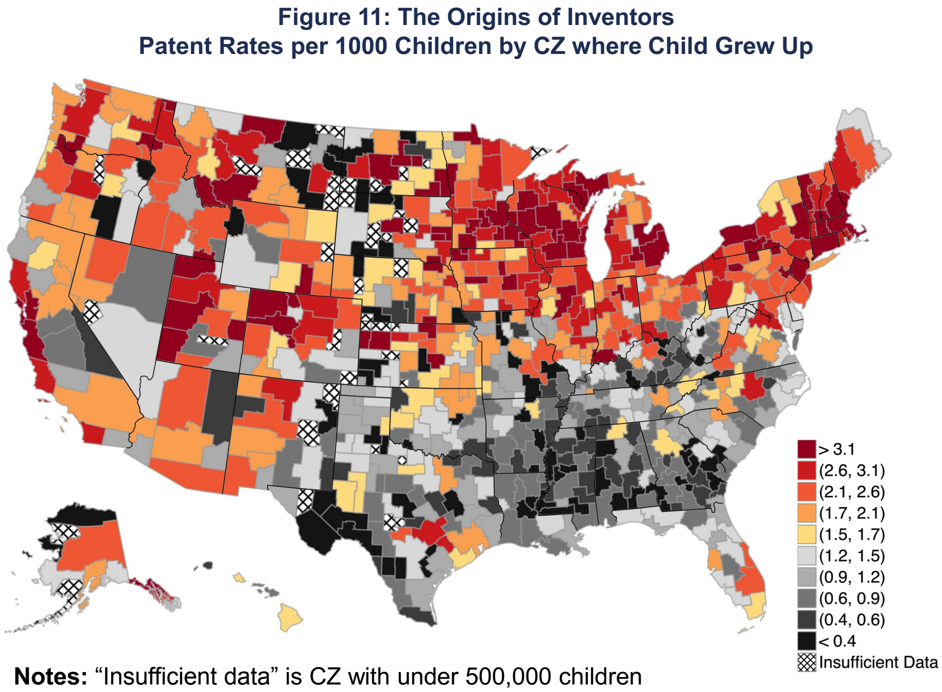
Our third measure of exposure looks at the geography of innovation. The existing literature examines the effects of place on economic outcomes in general and the effects on innovation in particular. Marshall's theory of industrial districts suggested that one advantage of geographical clusters was that there were "ideas in the air" and accounts of the success of Silicon Valley have also stressed the benefits of geographically localized innovation spillovers.[29] From our data we know where all individuals were grew up and in Figure 11 we present the invention rates by childhood Commuting Zone (CZ). Note that the figure does not present the patenting rates based on where inventors are *currently* residing (this is generally used in the literature because location at time of invention is available in the USPTO data), but rather the future patenting rates in commuting zones where inventors grew up. Figure 11 shows an interesting spatial pattern: innovation hotspots are in the Bay Area, Northeast and Great Lakes, which is expected as there are strong universities located there. But there are also many other innovation pockets around the country - in the North-West, Utah and Colorado, which may be less obvious hotspots. Very low innovation areas are found in the Deep South.[30]

[28] Figure A11 shows a bar chart of these industry regression coefficients grouped by distances. The falloff from innovating in the exact same class to the next closest class is very striking.

[29] There is a large literature on inventor mobility and agglomeration effects from localized knowledge spillovers, e.g. Kim & Marschke (2005).

[30] The list of the top 10 and bottom 10 most innovative CZs are in Table A5.

Figure 4.11.: The Geographic Origin of Inventors



These relationships cannot be read as causal place-based effects as there are many other unobservable factors associated with place of birth and future outcomes. We probe this relationship more in Table 4 which is analogous to Table 3 and again focus on children whose parents were not inventors. Column (1) is the baseline regression where the dependent variable is the fraction of kids who lived in a commuting zone that grow up to be inventors and the key right-hand-side variable is the invention rate in the childhood commuting zone (as in Figure 11). There is a strong and significant relationship between the two (as also illustrated in the scatterplot in Figure 12 for the largest 100 commuting zones) showing that children who grow up in innovation-intensive neighborhoods are more likely to become innovators themselves. According to column (1) of Table 4, increasing the fraction of inventors in the childhood commuting zone by a standard deviation (0.0002) is associated with a 30% increase in invention rates.[31]

[31] The mean of the dependent variable is 0.002, so using the coefficient estimate (0.0002*2.9006)/0.002 = 0.30.

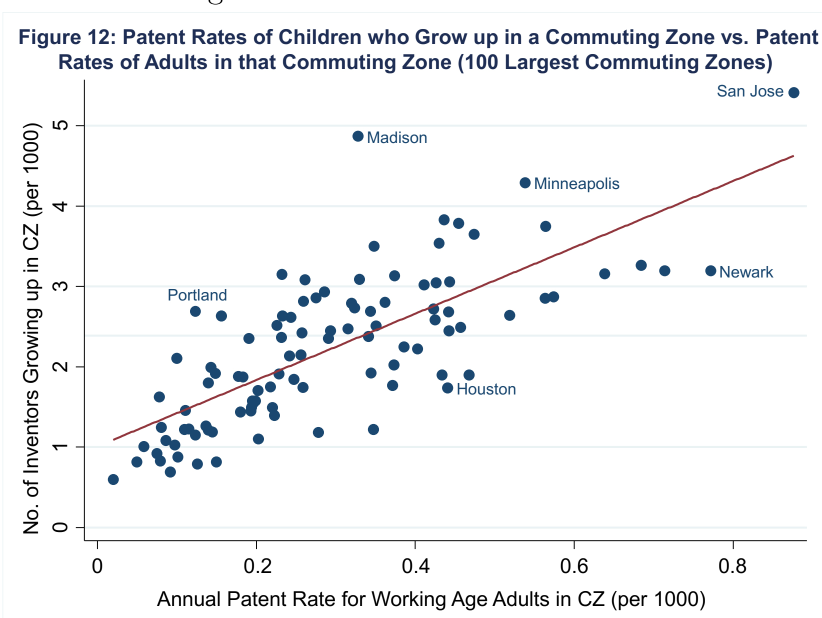
Table 4.4.: Children's Patent Rates vs. Patent Rates in Neighborhood

Table 4: Children's Patent Rates vs. Patent Rates in Neighborhood

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable: Fraction Children inventing	Fraction inventing in CZ	Dummy if individual invented	Fraction Inventing in in cell	Fraction Inventing in cell	Fraction Inventing in in cell	Fraction Inventing in cell	Fraction Inventing in cell	Fraction Inventing in cell
Parent's CZ patent rate	2.906*** (0.435)	3.098*** (0.572)						
Parent's CZ patent rate in Category			1.724*** (0.409)	1.378*** (0.396)	1.955*** (0.448)			
Parent's CZ patent rate in Sub-Category						1.509*** (0.386)		
Parent's CZ patent rate in Technology Class							1.136*** (0.194)	1.050*** (0.173)
Parent's CZ patent rate in same Sub-Category, other class								-0.0005 (0.0064)
Parent's CZ patent rate in same category but other Sub-Category								-0.0018 (0.0027)
Parent's CZ patent rate in other category								0.0053*** (0.0007)
Fixed effects	None	Child CZ	Category	Child CZ & Category	Parent NAIC & Category	Sub-Category	Technology Class	Technology Class
Cells	Parent CZ	Individual	Parent CZ*category	Child CZ*Parent CZ*Category	Parent CZ*Parent NAIC*Cat.	Parent CZ*Sub-category	Parent CZ*technology class	Parent CZ*technology class
Observations	390	5,452,642	2,730	1,554,973	1,642,193	14,430	173,550	173,550

Notes: Standard errors clustered by 390 Commuting Zones (CZ) where children grew up ("parent's CZ"). Columns (2) and (4) include current CZ as of 2012 ("child's CZ) fixed effects. Sample is children whose parents were not inventors.

Figure 4.12.: Patent Rates of Children Growing Up in a Commuting Zone vs. Patent Rates of Adults in that Commuting Zone



Column (2) of Table 4 reports estimates at the individual level and includes 390 dummy variables for the commuting zone where the individual was living in 2012. The coefficient is essentially identical to column (1). This is effectively exploiting the origin-destination CZ matrix of where a child grew up and where they currently live (almost 390*390 cells). The coefficient indicates that, for example, for two adults currently living in Chicago, the fact that one grew up in high-innovation Cambridge makes it more likely she will be an inventor than another who grew up in lower-innovation Little Rock. In column (3) we use CZ by patent category level (analogously to Table 4 column (2)) and include seven category fixed effects. Column (4) disaggregates this further by the CZ where the child grew up (so a cell is the current child's CZ by childhood parent CZ by category). Like column (2), this shows that there is a positive association with exposure even after controlling for the CZ where the child is currently living. Column (5) disaggregates the current child CZ by patent category and by father's industry. We can control for industry dummies here to show that the CZ exposure is picking up more than simply the parent industry exposure that we analyzed in Table 3. Column (6) is like column (3) but at the CZ by sub-category level (and includes sub-category fixed effects) and columns (7) and (8) are at the CZ by technology class level (regressions include class fixed effects). This table illustrates that not only are kids who grow up near inventors more likely to be inventors themselves, but these children also have an increased propensity to innovate in the same types of fine-level technologies as they were exposed to during childhood.[32]

4.4.2.4. Exposure to female inventors

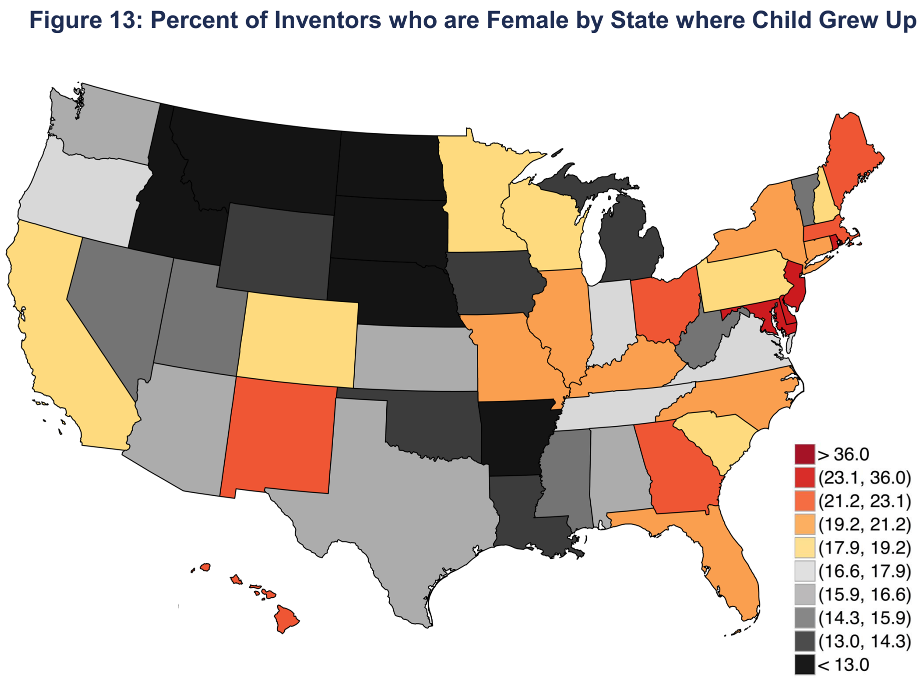
We can also use our empirical strategy of exposure to innovation in general to look more specifically at female inventors. We can calculate the proportion of female inventors in the state where children grew up and we present the map of this in Figure 13.[33] The Northeast appears to have the most female inventors and the North and mid-West the least. The top

[32] Figure A12 shows a bar chart of these CZ regression coefficients grouped by distance. The fall off from innovating in the exact same class to the next closest class is very striking

[33] Unfortunately, there are not sufficient numbers of female inventors to do this reliably at the CZ level

and bottom 10 CZ for female inventors are in Table A6. These female invention rates are correlated with the Pope & Sydnor (2010) Gender Stereotype Adherence Index on 8th Grade Tests (see Figure A13). It seems plausible that areas where these cultural attitudes about women are stronger are less likely to generate female inventors.

Figure 4.13.: Percent of Inventors who are Female by State where Child Grew Up



4.4.2.5. Summary of Findings on Exposure

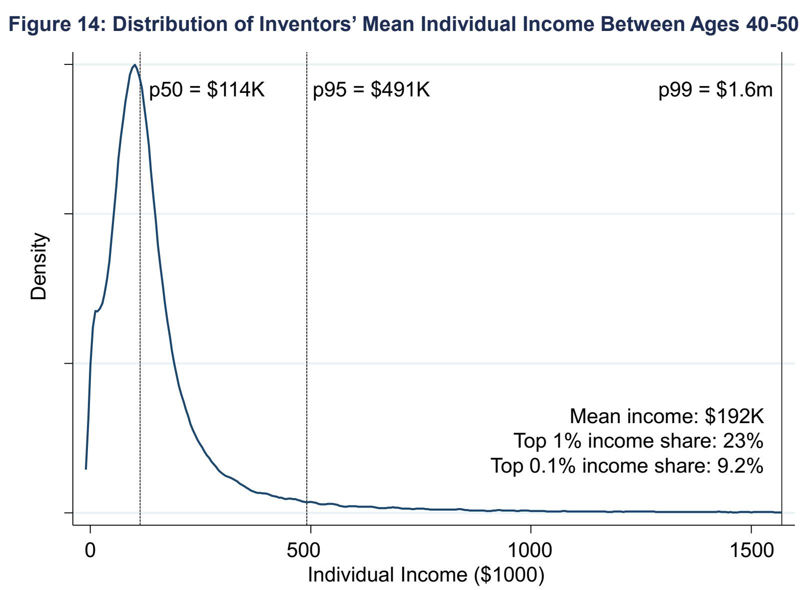
A clear pattern emerges from this section that across all three measures of exposure - parents, parents' colleagues and neighborhoods - growing up surrounded by more innovation is associated with an increased probability of becoming an inventor, even within a detailed technology class.

4.5. Inventors in the Labor Market

4.5.1. Wages and Income

The distribution of inventor income is given in Figure 14. It is heavily skewed with a median of \$114,000, a mean of \$192,000 and \$1.6m at the 99th percentile. It is far more skewed than occupations such as doctors, architects or lawyers. Only the financial sector has similar degrees of skewness (Lockwood et al., 2014). Examining the composition of income for inventors, we observe that it is dominated by salaries - like for most workers. Focusing on the top 1% of inventors by income, non-wage income is more important than wage income. However, even here patent royalty income is not so important: other forms of non-wage income dominate (Figure A14).

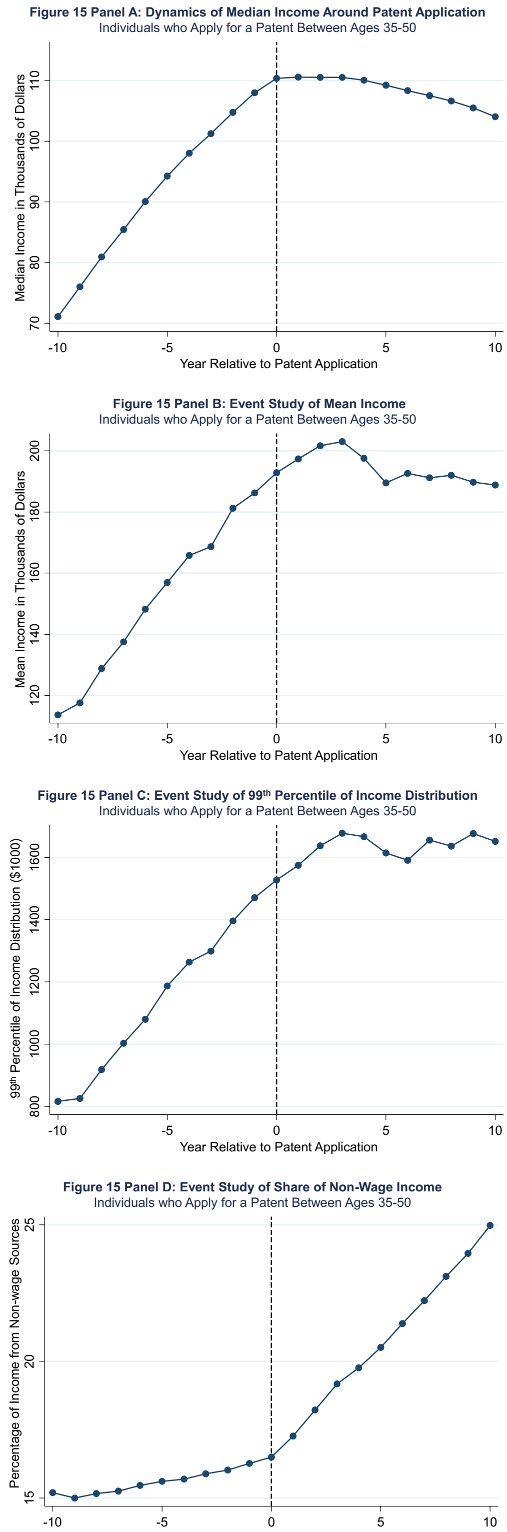
Figure 4.14.: Distribution of Inventors' Mean Individual Income Between Ages 40 and 50



Next, we examine income dynamics around the time of invention in Figure 15. Panel A shows median income in the 10 years before and 10 years after the patent event. Interestingly, we do not see an increase in income after applying for a patent. In fact, there is a much stronger increase in income prior to patenting: afterwards income flattens off and falls a bit. Panel B shows the pattern for mean wages and Panel C for the 99th percentile to check if the median is missing out on higher quantiles. But the picture looks similar with the main returns occurring prior to the patent event (although the flattening out now happens about three years after the patent is applied for). Panel D examines the proportion of income that is non-wage. Here we do see a clear break in trend with the fraction rising afterwards much faster than it was before. Panel E looks at alternative measures of the quality of patents. The patterns look similar, although ungranted appear less valuable than granted and highly cited patents (i.e. those in the top 5% of the future citation distribution by cohort have faster growth of income in the pre-patent period and the fall off of income in the post-patent period is less severe).[34]

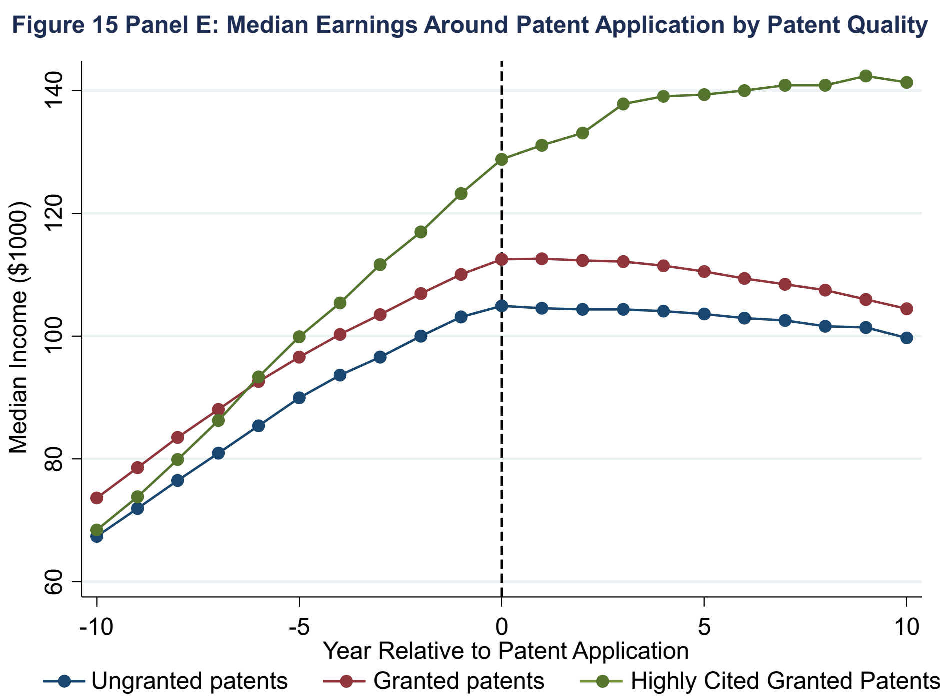
[34]Note that mean income is higher in general for inventors with more citations (see Figure A15). This is a new external corroboration of the value of citations as indicators of quality.

Figure 4.15.: Dynamics of Income Around Patent Application



The patterns in Figure 15 are for prime-aged workers (35-50 years old) to abstract away from the fast growth of earnings that occurs for younger workers. Figure 16 shows the same median income "event studies" for patents applied for at exactly the ages of 30, 40 and 50. The pattern of a fast growth prior to patenting followed by a leveling off/fall is apparent in these graphs as well, suggesting the dynamic patterns are occurring at all ages. It is not some conflation of lifecycle age-earning profiles with patent events.

Figure 4.16.: Distribution of Inventors' Mean Individual Income Between Ages 40 and 50



Our interpretation of the data is that there are some individuals who are on a successful innovation streak, culminating in applying for a patent. But the patent event is not news to the firm or labor market. The eventual patenter is obtaining rewards prior to patenting rather than afterwards. This is consistent with the evidence from matched patents-income administrative data from Italy in Depalo and Di Addario (2014) who also found big increases in income prior to patenting, but not afterwards.

These patterns suggest a model of career choice with an uncertain chance of a major innovation, rather than a choice about how and when to patent *per se*. The model in Section 6 builds on this insight.

4.5.2. Inventor Age

The most common age of patent applicants is approximately 39-40 (Figure A16 Panel A). However, this should be compared to the overall age distribution of US workers in the whole IRS population (Panel B). It is clear that most inventions tend to come from inventors who are older than the working population as a whole. Innovation still peaks around 40, but trails off more gradually than in Panel A (there is a 32% decline from 2.5 per 1,000 wage earners at age 40 to 1.71 at age 60). Does this over-estimate the abilities of the middle-aged? For example, could the patents of older workers be of lower quality and the most important work still be done when researchers are young? Panel C uses only highly cited patents and finds a similar pattern except with a sharper decline in innovation at older age (from 0.14 at age 40 to 0.06 at age 60: a decline of 57%).[35] Therefore, consistent with the "young and restless" hypothesis of Acemoglu *et al.* (2014) and the evidence in academia of "great" discoveries (e.g. Jones *et al.* (2014)) raw patents do under-estimate the importance of youthful innovation. However, there does seem to be a lot of innovation done at older ages too. Thus, the returns to innovation accrue to inventors relatively late during their career, which speaks to the importance of occupational choice early in life (extensive margin), as opposed to a choice "at the margin" to adjust one's innovative effort during one's career (intensive margin).

[35] Figure A16 shows that overall patent citation rates are correlated with individual income which is another corroboration of the idea that citations are a useful measure of patent value.

4.5.3. Firms

We know what firms inventors work in from the EIN number attached to their W2s. Corporate structures are complex, however - for example, a US parent company may have multiple affiliates, the ultimate headquarters may be in another country and there are partnership structures as well as firms. We simply work with the EIN numbers and discuss some of the data issues in Appendix A.[36] Figure A17 has the CDF of the size distribution for inventors. 10% of inventors do not receive a W2, presumably because they are self-employed. But the bulk of inventors work and mostly they work for larger firms. 70% of inventors work in firms with more than 100 employees. Hence, it is unlikely that the pattern in Figure A17 could be due to individuals simply being unable to afford the fees associated with applying for a patent. It is the companies that inventors work for who are applying and paying for these fees. And although in a small firm the inventor himself may have to contribute, this is highly unlikely to be the case in a larger firm.

4.5.4. Summary

In summary, the key findings from our labor market analysis are (i) there are substantial returns before the patenting itself; (ii) patenting returns appear very skewed: there is a small chance of a very high payoff (cf. Hall and Woodward, 2010, on entrepreneurs) and (iii) many returns are late in an inventor's career (cf. Jones, 2009, 2010) . Hence pay-offs are highly uncertain when individuals are making initial career choices.

[36] The distribution of employment by firm size class in the IRS dataset, where each firm is assumed to be represented by a unique EIN, is almost identical to the size distribution across the US economy from the Economic Census. Despite many reasons why the Census Bureau concept of an "enterprise" could be quite different from the tax-based EIN, the match with the firm size distribution is very good. There appears to be a slight under-representation of firms with under 100 employees in the IRS data compared to the Census, and a slight over-representation in the 100-1000 employee range, but almost identical proportions in other size bins compared to the Census. And even the discrepancies are small. Narrowing the IRS sample to be closer to the Census in terms of industry composition does not fundamentally change this picture.

4.6. A Simple Model of Inventors' Careers

4.6.1. Basic Model

Motivated by the stylized facts from our analysis we develop an inventor lifecycle model that is detailed in Appendix B. The model has similarities to a two-sector version of Hsieh et al. (2013). In period 1 human capital (H) is determined, which will depend on initial talent (which is heterogeneous across individuals and groups) and schooling.[37] At the end of schooling, agents make occupational choices over whether to enter the R&D sector or non-R&D sector on the basis of their expected utility. The R&D sector is different from the non-R&D sector in three respects: (i) Income is stochastic: there is a base wage (as in the non-R&D sector) but there is also a chance (π) of making a successful innovation; (ii) those with high human capital have a comparative advantage in the R&D sector - formally, we model this as allowing the probability of innovating to be higher when human capital is higher ($\pi(H), \pi'(H) > 0$); (iii) individuals have idiosyncratic preferences over the two sectors (hence, there will typically be some mass of agents who go into the R&D sector even if their expected monetary returns are lower). There is a two-part tax regime with a standard marginal tax rate up to a threshold and then a high "top" tax rate.

We consider that there are several disadvantaged groups, g, in the population. The canonical example we focus on are children born to low income families (so the groups are rich vs. poor), but the framework is equally applicable to considering men vs. women or whites vs. blacks. We allow for these groups to face additional costs in the acquisition of human capital (which reduces their likelihood of entering the R&D sector).[38]

[37] In the baseline model we keep the human capital acquisition process exogenous (e.g. high income families "buy" more educational quality for their children). But we extend the model to allow for endogenous educational choice. This reinforces the misallocation results as disadvantaged groups choose to invest less in human capital as they are less likely to want an inventor career.

[38] Hsieh et al. (2013) model this as a "tax friction" on spending goods in obtaining human capital. Another friction is direct discrimination in the labor market by paying disadvantaged groups a lower wage than their

We depart from the standard rational sorting models by adding imperfect information. We assume that individuals with probability λ are correctly informed about the net returns to an inventor career, but others (with probability $1 - \lambda$) under-estimate the true returns from the R&D sector. The idea is that many children do not often come into contact with inventors via their parents, family networks or neighborhoods. Hence these less "exposed" groups underestimate the net benefits of choosing an inventor career (e.g. Hoxby and Turner, 2014). The precise way we formalize this notion is that these less-informed agents underestimate the degree of complementarity between human capital and innovation. We show a number of intuitive results using this model.

First, agents with higher human capital are more likely to enter the R&D sector (this follows directly from their comparative advantage). Second, individuals who are more exposed to inventors (our proxy for a higher value of λ) are more likely to enter the R&D sector. Third, disadvantaged groups are less likely to enter the R&D sector. This is because (i) they face barriers to human capital acquisition and (ii) they may begin with a lower ability level (for example, we saw the 3rd grade math scores of low income groups were lower), and (iii) they have worse information about the net benefits of an inventor career.

A fourth result is that conditional on being in the R&D sector, agents from disadvantaged groups are likely to have lower levels of initial ability. This is the opposite prediction from the standard rational sorting models which predict higher levels of average ability for disadvantaged groups in the R&D sector, because only the most talented will overcome the barriers to entry. By contrast, in this model agents from disadvantaged groups are imperfectly informed about the complementarity between human capital and innovation. Therefore, the high ability agents in these groups do not always go into the R&D sector, whereas for privileged groups the high ability agents do.

---

marginal product (in the R&D sector compared to the non-R&D sector). This is less likely for low income groups, but could be the case for women or minorities. In the context of their model these are observationally equivalent.

The empirical evidence lines up well with these predictions. We find that people with higher human capital (test scores and elite college attendance) are more likely to be inventors, and that disadvantaged groups are less likely to become inventors. Section 4 presented much evidence that early exposure mattered a lot for future innovation. Finally, Figure 8 showed that conditional on being an inventor, disadvantaged groups did not appear more talented than other groups. This is consistent with our imperfect information model but not with the basic rational sorting model

The normative implications of our model is that there is the potential for considerable welfare loss. First, there is a "level effect" - too few individuals from disadvantaged backgrounds enter the R&D sector. This effect is similar to Hsieh et al. (2013). Second, and in contrast with standard rational sorting models, there is also a "composition" effect: the "wrong" individuals from disadvantaged groups may enter the R&D sector due to information frictions (i.e. not the individuals with the highest comparative advantage for the R&D sector). And these effects are compounded because the disadvantaged groups will likely face other barriers. There is a loss both from fewer (externality generating) inventors and from an inferior allocation of talent compared to the first best.

4.6.2. Alternative Interpretations of Exposure Measures

We have interpreted our empirical measures in terms of exposure influencing λ. But one alternative is that the exposure measures actually improve inventor-specific human capital as the young person grows up. This would be an environmental effect, but the welfare effects are somewhat different from the model of the previous section. At the time of occupational choice, those exposed in childhood would have higher human capital rather than more information (λ). Hence, there would not be obviously greater talent losses in this model than in the rational sorting model.

As a way to investigate this we examine the income effects of exposure. If exposure generates inventor-specific human capital then we would expect productivity (and therefore wages) to be higher for those choosing an inventor career if they are more exposed at an earlier age to science, even conditional on early test scores. However, we could find no evidence of substantially higher wages for these groups more exposed to innovation. In our model the effects of exposure on inventor income are ambiguous. Although agents are better sorted to their area of comparative advantage as information improves, they may earn a lower wage in the R&D sector as there are compensating differentials to the non-pecuniary advantages of the R&D sector ($\widetilde{w}$). [39]

Another interpretation of the exposure "effect" is that it could cause a change in preferences rather than in information. The wage effects of this are also ambiguous, so it is difficult to empirically distinguish this from our information story. In terms of policy, if there are externalities from innovation then increasing exposure to inventors might still be highly valuable even if it is only about changing preferences. Note that the pure preference shift story would not predict that exposure should be more important for high-ability children in their likelihood of growing up to be an inventor. It is hard to check this directly in our data because there is insufficient variation in exposure in the NYC data to include interactions between exposure and early test scores in the "inventor equation"[40], but this would be a good avenue for future work.

There are, of course, other ways of interpreting our results but overall, the simple lifecycle model we sketch seems reasonably consistent with the data.

[39] For example, Stern (2004) shows that life science post-graduates take about a 20% loss in income by taking a job in academic science compared to industry.

[40] A positive coefficient on the interaction would be consistent with our model, but not with the pure preference model.

4.6.3. Policies to improve entry into innovation

If policy interventions could improve the position of potentially high-ability kids from poor backgrounds (as well as women and minorities) this would bring a whole new margin of individuals into the inventor pool. Card *et al.* (2013) report evidence from "gifted and talented" randomized control trials that suggest that although these programs do not work well for the typical student, those from poorer backgrounds do appear to particularly benefit. Changing to such policies has a near-zero financial cost.[41] This suggests that such interventions could have very large benefits in terms of growth as well as equity.

Our estimates can be used to assess the potential gains from such supply-side educational policies. These policies can have a *level* effect on innovation by increasing the rate of entry into innovation of children from low-income families, up to a level closer to that of children from high-income families. In addition, the policies may have a *composition* effect by affecting which children decide to enter the R&D sector within each income group - our results suggest that talented children from low-income families are less likely to enter the R&D sector, which could potentially be affected by policy). Appendix B4 discusses these calibrations in more detail, but we sketch the findings briefly here.

Regarding the level effect, we have shown that children born to families in the top 10% of income are ten times more likely to become inventors than children born to families of below-median income (Figure 1). We have documented that innate ability differences (which by definition cannot be affected by policy) are unlikely to explain more than a third of this difference (Table 1). Moreover, we have found that the differential innovation rates across technology classes is also a ten-to-one ratio (Figure 10), which suggest that exposure effects play a key role.

We consider a benchmark scenario assuming that supply-side policies providing such expo-

[41] Personal communication with David Card.

sure effects could close a fifth of the total innovation gap between children with parents below the 90th percentile of income (1.6 inventors per 1,000) and children with parents in the top 10% (6.7 inventors per 1,000). Under these assumptions and using our data on the propensity of children to become inventors across the income distribution, the increase in the number of inventors induced by the policy is a staggering 30% of the current inventor population. The details of the calculation are reported in Appendix B5, where we also report the calibrated effect of the policy under other assumptions about the share of the innovation gap across the income distribution that can be closed by policy. The sensitivity analysis shows that the effect is large under a very wide set of parameter values. We have also considered a scenario based on the distribution of innovation-income gaps across US states.[42] If the mean innovation-income gap (similar to Michigan) could be lowered to the level of the fifth percentile of the distribution (similar to states like New Hampshire), then the increase in the number of inventors would be equal to 38% of the current inventor population. Similar calibrations suggest the composition effect also matters, but by less than the levels effect (under 6%).

By contrast to these "extensive margin" policies, policies on the intensive margin may be much less effective. For example, although existing estimates of R&D tax credits do suggest increases in innovative activity, they reach only a small fraction of the population. We examine the effects of income tax on innovation next.

4.6.4. How much do tax policies on inventor's income affect innovation?

A much-discussed alternative set of policies to stimulate innovation are levels of top income tax rates (Akcigit *et al.* (2015)). We shed light on this issue using a quantitative theoretical

[42] All calculations are based on state of birth, not on state of residence.

exercise based on our data and the lifecycle model introduced in sub-section VI.A (more details are in Appendix B.3). In short, we find that changing marginal tax rates on high incomes would *not* substantially affect the occupational choice to become an inventor. The small effect of top income marginal tax rates on inventors' behavior is driven by three considerations. First, marginal utility decreases with income, due to risk aversion. Second, inventors' earnings are highly skewed - as documented in Section V. Third, the rank of an inventor in the earnings distribution has a large random element, especially within the top tail. We find that income is difficult to predict for inventors, especially in the upper tail, therefore we model it as a random draw from the empirical income distribution for all inventors.[43]

Put another way, the intuition for the results is that entering the innovation sector is like buying a lottery ticket. With concave utility and expected utility maximization, whether an agent has a small chance of winning $20 million (e.g. in an economy with no extra taxation of top incomes) or an equally small chance of winning $10 million (e.g. in an economy with a higher marginal tax rate of top incomes) makes little difference to buying the lottery ticket. In other words, occupational choice over whether to become an inventor should not respond much to top tax rates. Our contribution is to calibrate this response using the skewness of the empirical earnings distribution of inventors. It is important to note that our approach does not say anything about the broader effect of top tax rates on the rate of innovation in

[43] This is a strong assumption, but various features of the data support its validity to a first approximation. First, we have shown in Section V that the returns to patenting are very skewed and that they often occur late in an inventor's career, suggesting that the returns to innovation are very uncertain at the time of occupational choice. Second, the distribution of income for inventors is very skewed even conditional on having a very highly cited patent: the 99th percentile of income for people in the top 5% in the citation distribution is $10 million, while the median is about $200,000. In other words, even in the subgroup of inventors who produced high-quality innovations, the mean return is coming almost entirely from the upper tail. Third, we have checked that an inventor's earnings in mid-career, between ages 40 and 45, are very difficult to predict based on this inventor's earnings and patenting record (number of patents and citations) in his early career, between ages 27 and 32. Specifically, regressing mean earnings between ages 40 and 45 on a flexible polynomial of mean earnings, patent count and citation-weighted patent count between ages 27 and 32 yields a R^2 below 0.03.

the economy.[44]

For the calibration, we consider a variety of tax regimes. We use a stylized version of the US Federal tax schedule where the top marginal tax rate is 40% above \$439,000 and the marginal tax rate below this threshold is 28.5% ("standard rate"). We then consider the impact on innovation of increasing the top rate by a percentage point to 41%, but keeping the standard rate the same at 28.5%. Since the benefit of the policy is to raise revenue for public goods, we have to benchmark this in some way. So we consider a "benchmark" policy of raising the standard rate by a percentage point (to 29.5%) but keeping the top rate the same. We then calculate the fall in the fraction of the population of workers becoming inventors per dollar of tax revenue in both cases.[45] This is equivalent to a (marginal) deadweight cost per tax dollar. We denote the loss of inventors per tax dollar due to the higher top rate policy as $\gamma(\tau^1)$ and the corresponding deadweight for the benchmark policy of raising the standard rate as $\gamma(\tau^B)$. We then repeat these calibrations under various assumptions about the utility function.[46]

As is standard in public finance, the deadweight cost crucially depends on a behavioral elasticity, which in our context captures the extent to which number of people choosing an inventor career responds to the change in the certainty equivalent wage induced by changes in the tax system. However, this elasticity cancels out when we express the relative innovation loss of any policy change relative to the benchmark policy change described above. In other words, we focus on a summary statistic $\gamma = \frac{\gamma(\tau^1)}{\gamma(\tau^B)}$, which we can calibrate based only on the empirical income distribution of inventors, without knowledge of the behavioral elasticity,

[44] Indeed, beyond inventors, many other agents are involved in the innovation process, for instance firms and financiers, for whom the returns to innovation may not be analogous to a random draw (e.g. because they hold large and diversified portfolios of innovations). Moreover, our analysis does not take general equilibrium effects into account.

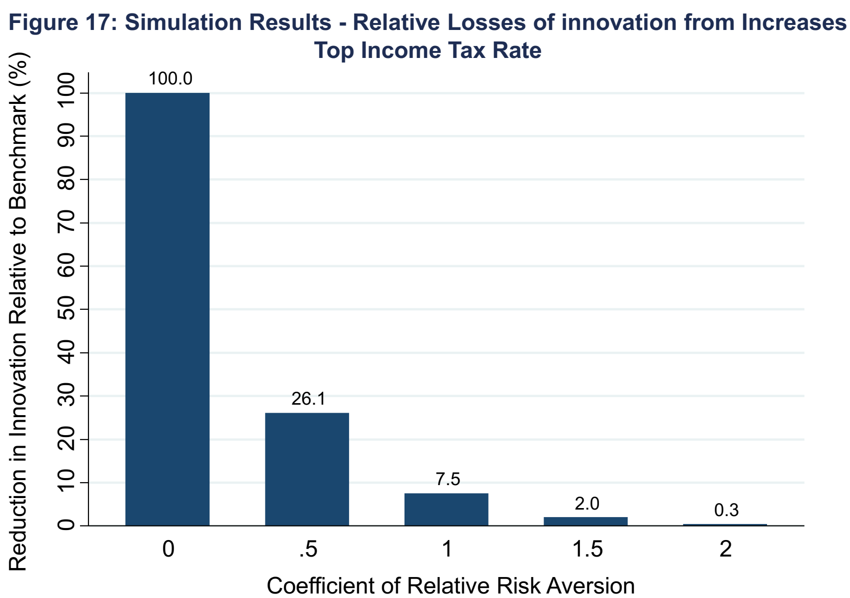
[45] We compute the fall in the fraction of the population of workers becoming inventors based on the change in the certainty equivalent implied by changes in the tax system. Appendix B describes this and all other steps of the calibration in detail.

[46] Specifically, we consider CRRA utility functions with coefficients of relative risk aversion equal to 0 (i.e linear utility), 0.5, 1 (i.e. log utility), 1.5 and 2, respectively.

and which captures the relative efficiency loss from increasing the top marginal tax rate compared to our policy benchmark of the change in the standard tax rate (Appendix B.3.1 provides a full derivation of these results).

We present the value of γ in Figure 17 for different levels of the coefficient of relative risk aversion in the utility function. The first bar is 100% for a CRRA parameter equal to 0 (linear utility): when people are risk neutral, there is no difference in changes in top and standard tax rates. As we move to the right in Figure 17, we consider increasing levels of risk aversion. The height of the bars falls indicating the welfare loss from raising top taxes is diminishing as risk aversion increases. It is about a quarter of the innovation loss of the benchmark policy when CRRA = 0.5 and only 2% of the benchmark policy when CRRA = 1.5.[47]

Figure 4.17.: Distribution of Inventors' Mean Individual Income Between Ages 40 and 50



Notes: These are the losses in the fraction of inventors from changing the rate of top income taxes by 1% compared to a benchmark policy of increasing the standard rate of tax by 1%. The x-axis compares three different levels of risk-aversion (parameter δ in the Constant Relative Risk Aversion (CRRA) utility function, $u(c) = c^{1-\delta}/(1-\delta)$ where c = consumption). The graphs show that for plausible levels of risk aversion, changes in fraction of inventors is small The details of the model and simulation are in Appendix B.

[47] Appendix B shows some other illustrations of this effect when looking at larger changes in top taxes and additional thresholds. We show that the skewness of the earnings distribution drives our quantitative results: in the presence of concave utility and stochastic payoffs, taxes have a lower efficiency cost for very skewed payoff distributions compared with less skewed ones.

The intuition behind the result is clear. The innovation loss from high top tax rates is to discourage entry into the riskier R&D sector - if an inventor is lucky and wins a very valuable innovation, she will be penalized more by high tax rates. But since the returns to innovation are so skewed, most inventors obtain only low returns. There can be considerable differences in income in the tails, but the utility difference will be much less due to the concavity of the utility function.

The generality of this result should not be overstated. It is a theoretical result based on calibrated parameters in a standard expected utility set-up, which crucially depends on the assumption that the income process for inventors is as good as random. If inventors are not fully rational expected utility maximizers, they may well behave differently. Furthermore, even in the context of the model it may be that potential inventors are better at predicting their expected income than we assume. If there was no uncertainty about the returns to innovation, then the results would no longer hold.[48]Finally, our evidence does not shed light on broader effects of changes in tax rates on innovation dynamics.[49]

While these caveats must be kept in mind, our point is still an important one: we show that the effect of top tax rates on the key individuals in the innovation process - the inventors themselves - is likely to be small due to the skewness and randomness of the payoffs.

[48]There is a recent empirical literature on the impact of top tax rates on inventor mobility. Akcigit *et al.* (2015) find that differences in top marginal tax rates across countries effect mobility decisions of inventors. However, whilst statistically significant the magnitude of the effect is quantitatively small in line with our calibrations. Moretti and Wilson (2015) find larger effects when looking at inventor mobility across states in response to changes in top income tax rates. This within US analysis is interesting but it does not speak to the issue of whether people are more likely to choose an inventor career, as the empirics work from mobility once people have already chosen to become inventors.

[49]In particular, it could be the case that the returns to innovation are much less skewed for a number of agents who play a crucial role in the innovation process and hold a large and diversified portfolio of innovations, for instance investors (venture capital firms, etc.) and R&D lab managers or CEOs. In other words, it could be that an increase in top tax rates would lead these agents to exert less effort at the margin, because there is a tight link between their effort level and their financial rewards (in contrast with what the data suggests for inventors themselves). For example, Aghion *et al.* (2015a) suggest that the returns to innovation are broadly shared, much beyond inventors, and that innovation has contributed to the increase in top income inequality in recent decades.

4.7. Conclusion

In this paper we describe the lifecycle of inventors using a match between 1.2m patentors during 1996-2014 with administrative tax data from the IRS. We are able to follow potential inventors from their conditions at birth (parental income, gender and race), while growing up (neighborhood, school text scores, college attended) and finally in the labor market (their income profiles). We show that parental characteristics matter a lot: the rich, white and male are much more likely to grow up to be inventors than the poor, female and black. Early math test scores account for a third of the inventor gap for income, a tenth for the black-white difference and almost nothing for gender. Focusing on the differences by parental income, two elements prove to be key. First, the innovation-relevant human capital gap between rich and poor opens up as they go through school, and by the time they reach college parental income makes relatively little difference. Second, children who are most exposed to innovation through their parents, mentors or neighborhoods are much more likely to become inventors.

These findings suggest there are environmental barriers to disadvantaged groups in acquiring the human capital that is complementary with innovation. The predictions of standard “rational sorting” models that inventors from disadvantaged groups should be of higher quality (in terms of patent citations and early ability) is not born out by the data. We discuss an extension to the sorting model of inventor careers in which some individuals underestimate the benefits of a career as an inventor through less exposure and show that its predictions are broadly consistent with the data. Policies that reduce the income-related test score divergence (particularly for children in the top decile of early math achievement) are beneficial from an equity stance, but might also be beneficial from an efficiency perspective as they have the potential to increase the degree of innovation and growth in the US economy. Through a series of calibrations, we have shown that such “extensive margin” innovation policies drawing talented low-income individuals into innovation may be very effective at stimulating innovation, probably more than top income tax policies.

There are many more aspects of the data we have discussed here that could be used to inform the next generation of innovation models. How important is entrepreneurship in enabling innovators to successfully monetize high-quality inventions? Can we bring more structural models to bear on the inventor career model in order to consider counter-factual policies? How important are firm policies in better attracting and incentivizing inventors? And could the same forces which influence whether a person becomes an inventor hold true for other elite occupations? These are some of the fascinating questions our research agenda opens up for the next generation of innovation theories and policies.

Bibliography

Abadie, Alberto, & Imbens, Guido W. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**(1), 235–267.

Abadie, Alberto, & Imbens, Guido W. 2008. On the failure of the bootstrap for matching estimators. *Econometrica*, **76**(6), 1537–1557.

Acemoglu, D., Akcigit, U., & Celik, M. 2014. Young, restless and creative: Openness to disruption and creative innovations. *Nber working paper no. 19894*.

Acemoglu, Daron. 1998. Why do new technologies complement skills? directed technical change and wage inequality. *Quarterly journal of economics*, 1055–1089.

Acemoglu, Daron. 2002. Directed technical change. *The review of economic studies*, **69**(4), 781–809.

Acemoglu, Daron. 2007. Equilibrium bias of technology. *Econometrica*, **75**(5), 1371–1409.

Acemoglu, Daron, & Autor, David. 2011. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, **4**, 1043–1171.

Acemoglu, Daron, & Linn, Joshua. 2004. Market size in innovation: Theory and evidence from the pharmaceutical industry. *Quaterly journal of economics*, 1049–1090.

Acemoglu, Daron, & Robinson, James. 2012. *Why nations fail*. Crown Publishing Group.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, Peter. 2005. Competition and innovation: An inverted u relationship? *Quarterly journal of economics, 120(2), 701-728*.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., & Hemous, D. 2015a. Innovation and top income inequality. *Lse mimeo*.

Aghion, P., Akcigit, U., Hyytinen, A., & Toivanen, O. 2015b. Living the american dream in finland? *Lse mimeo*.

Akcigit, U., & Kerr, W. 2013. Growth through heterogeneous innovations. *Bank of finland discussion paper no. 1328*.

Akcigit, U., Baslandze, S., & Stantcheva, S. 2015. Taxation and the international mobility of inventors. *Nber working paper no. 21024*.

Argente, David, & Lee, Munseob. 2015. Cost of living inequality during the great recession. *Available at ssrn 2567357*.

Autor, David, Dorn, David, & Hanson, Gordon. 2013. The china syndrome: Local labor market effects of import competition in the united states. *The american economic review*, **103**(6), 2121–2168.

Azoulay, P., J.S., Graff Zivin, & G., Manso. 2010a. Incentives and creativity: Evidence from academic life. *Rand journal of economics.*

Azoulay, P., Graff Zivin, J. S., & Wang, J. 2010b. Superstar extinction. *Quarterly journal of economics, 125 (2), 549-589.*

Bessen, James, & Meurer, Michael J. 2013. Direct costs from npe disputes, the. *Cornell l. rev.*, **99**, 387.

Bloom, N., Schankerman, M., & Van Reenen, J. 2013. Technology spillovers and product market rivalry. *Econometrica (2013) 81 (4) 1347–1393.*

Bloom, N., Romer, P., S., Terry, & J., Van Reenen. 2014. A trapped factors model of innovation. *Nber discussion paper http://cep.lse.ac.uk/pubs/download/dp1261.pdf.*

Blundell, R, Griffith, R., & Van Reenen, J. 1999. Market shares, market value and innovation in a panel of british manufacturing firms. *Review of economic studies, 66, 529-554.*

Borjas, George J, & Doran, Kirk B. 2012. The collapse of the soviet union and the productivity of american mathematicians. *The quarterly journal of economics*, **1143**, 1203.

Borjas, George J, & Doran, Kirk B. 2015. Which peers matter? the relative impacts of collaborators, colleagues, and competitors. *Review of economics and statistics*, **97**(5), 1104–1117.

Breschi, S., & Lissoni, F. 2005. Cross-firm inventors and social networks: localized knowledge spillovers revisited. *Annales d'economie et de statistique 79/80, 189-209.*

Breschi, S., & Lissoni, F. 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of economic geography 9(4), 439–468.*

Broda, Christian, & Romalis, John. 2009. The welfare implications of rising price dispersion. *University of chicago mimeo*, **3**.

Broda, Christian, & Weinstein, David E. 2006. Globalization and the gains from variety. *The quarterly journal of economics*, **121**(2), 541–585.

Broda, Christian, & Weinstein, David E. 2010. Product creation and destruction: Evidence and price implications. *The american economic review*, **100**(3), 691–723.

Card, D., & Giulano, L. 2014. "does gifted education work? for which students ?". *Nber working paper 20453.*

Card, D., Heining, J., & Kline, P. 2013. Workplace heterogeneity and the rise of west german wage inequality. *The quarterly journal of economics, 128(3): 967-1015.*

Celik, M. 2014. Does the cream always rise to the top ? the missallocation of talent and innovation. *University of pennsylvania, mimeo.*

Chetty, R., Friedman, J. N., & Rockoff, J. E. 2014a. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review*, **104**(9), 2593–2632.

Chetty, R., Hendren, N., Kline, P., & Saez, E. 2014b. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *Quarterly journal of economics*, Jan.

Chetty, Raj, & Hendren, Nathaniel. 2015. The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. *Harvard university working paper*.

Chetty, Raj, Friedman, John N, Rockoff, Jonah E, *et al.* 2014c. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, **104**(9), 2633–79.

Cockburn, Iain, Kortum, Samual, & Stern, Scott. 2003. Are all patent examiners created equal? examiners, patent characteristics and litigation outcomes. *Intellectual property in the knowledge-based economy. national academy of sciences*.

Cohen, Lauren, Gurun, Umit, & Kominers, Scott Duke. 2014. *Patent trolls: Evidence from targeted firms*. Tech. rept. National Bureau of Economic Research.

Comin, Diego A, Lashkari, Danial, & Mestieri, Martí. 2015. *Structural change with long-run income and price effects*. Tech. rept. National Bureau of Economic Research.

Cook, L., & Kongcharoen, C. 2010. The idea gap in pink and black. *Nber working paper 16331*.

Cotropia, Christopher A, Kesan, Jay P, & Schwartz, David L. 2014. Unpacking patent assertion entities (paes). *Minn. l. rev.*, **99**, 649.

Depalo, D., & Di Addario, S. 2015. Inventors' returns to patents. *Bank of italy mimeo*.

Diamond, Rebecca. 2015. The determinants and welfare implications of us workers diverging location choices by skill: 1980-2000. *Working paper*.

DiNardo, J., Fortin, N. M., & Lemieux, T. 1996. "labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach". *Econometrica, 64(5), 1001-44*.

Ding, W., Murray, F., & Stuart, T. 2006. Gender differences in patenting in the academic life sciences. *Science*, **313**, 665–556.

Dorner, M., Harhoff, D., & Hoisl, K. 2014. Patterns and determinants of inventor mobility – evidence from the employment biographies of inventors in germany. *mimeo*.

Erickson, Tim, & Pakes, Ariel. 2011. An experimental component index for the cpi: From annual computer data to monthly data on other goods. *The american economic review*, **101**(5), 1707–1738.

Faber, Benjamin, & Fally, Thibault. 2015. Firm heterogeneity in consumption baskets: Evidence from home and store scanner data. *Working paper*.

Fallick, B., Fleischman, C. A., & Rebitzer, J. B. 2006. Job-hopping in silicon valley: some evidence concerning the microfoundations of a hightechnology cluster. *The review of economics and statistics 88, 472–481*.

Feenstra, Robert C. 1994. New product varieties and the measurement of international prices. *The american economic review*, 157–177.

Fernandez-Zubieta, A., Geuna, A., & Lawson, C. 2013. Researchers' mobility and its impact on scientific productivity. *University of turin working paper no. 13/2013*.

Frakes, Michael D, & Wasserman, Melissa F. 2014. *Is the time allocated to review patent applications inducing examiners to grant invalid patents?: Evidence from micro-level application data*. Tech. rept. National Bureau of Economic Research.

Franco, A. M., & Mitchell, M. F. 2008. Covenants not to compete, labor mobility, and industry dynamics. *Journal of economics and management strategy 17, 581–606.*

Fryer, R. 2014. Injecting charter school best practices into traditional publicschools:evidence from field experiments. *Quarterly journal of economics, 129(3), 1355-1407.*

Giuri, Paola, Mariani, Myriam, Brusoni, Stefano, Crespi, Gustavo, Francoz, Dominique, Gambardella, Alfonso, Garcia-Fontes, Walter, & Geuna, Aldo. 2007. Inventors and invention processes in europe: Results from the patval-eu survey. *Research policy 36(8),1107-1127.*

Goolsbee, A. 1998. Does government r&d policy mainly benefit scientists and engineers? *American economic review, 88(2), 298-302.*

Graham, Stuart, & Hegde, Deepak. 2015. Disclosing patents' secrets. *Science*, **347**(6219), 236–237.

Griliches, Zvi. 1992. The search for r&d spillovers. *Scandinavian journal of economics, 94, s29-47.*

Grossman, G., & Helpman, E. 1991. Quality ladders and product cycles. *Quarterly journal of economics, 106: 557-586.*

Hall, B., & Lerner, J. 2009. The financing of r&d and innovation. *Nber working papers 15325.*

Hall, B., Jaffe, A., & Trajtenberg, M. 2001a. The nber patent citation data file: Lessons, insights and methodological tools. *Nber wp 8498.*

Hall, B., Jaffe, A., & Trajtenberg, M. 2005. Market value and patent citations: A first look. *Journal of economics, 36, 16-38.*

Hall, Bronwyn H, Jaffe, Adam B, & Trajtenberg, Manuel. 2001b. *The nber patent citation data file: Lessons, insights and methodological tools*. Tech. rept. National Bureau of Economic Research.

Handbury, Jessie. 2013. Are poor cities cheap for everyone? non-homotheticity and the cost of living across us cities. *The wharton school research paper.*

Hoisl, K. 2007. Tracing mobile inventors - the causality between inventor mobility and inventor productivity. *Research policy, 36: 619-636.*

Hoxby, C., & Turner, S. 2015. What high-achieving low-income students know about college. *American economic review, 105(5) 514-17.*

Hsieh, C., & Klenow, P. 2009. Misallocation and manufacturing tfp in china and india. *Quarterly journal of economics, 124(4), 1403-1448.*

Hsieh, C., Hurst E. Jones C., & Klenow, P. 2013. The allocation of talent and u.s economic growth. *Mimeo, stanford.*

Hunt, J. 2009. Why don't women patent? *Nber working paper no. 17888.*

Jaffe, A., & Trajtenberg, M. 2002. *Citations and innovations.*

Jaffe, Adam B, & Jones, Benjamin F. 2015. *The changing frontier: Rethinking science and innovation policy.* University of Chicago Press.

Jaravel, Xavier, Petkova, Neviana, & Bell, Alex. 2015 (July). *Team-Specific Capital and Innovation.* SSRN Scholarly Paper ID 2669060. Social Science Research Network, Rochester, NY.

Jones, Benjamin F. 2010. Age and great invention. *Review of economics and statistics.*

Jones, B.F., Reedy, E. J., & Weinberg, B.A. 2014. Age and scientific genius. *Nber working paper no. w19866.*

Jones, C., & Williams, J. 1999. Too much of a good thing? *Journal of economic growth 5(1) 65-85.*

Jung, T., & Ejermo, O. 2014. Demographic patterns and trends in patenting: Gender, age and education of inventors. *Technological forecasting and social change 86, 110-124.*

Kahneman, D., & Ghiselli, E. 1962. Validity and nonlinear heteroscedastic models. *Personnel psychology (15) 1-11.*

Kane, Thomas J, & Staiger, Douglas O. 2008. *Estimating teacher impacts on student achievement: An experimental evaluation.* Tech. rept. National Bureau of Economic Research.

Katz, Lawrence F, Krueger, Alan B, *et al.* 1998. Computing inequality: Have computers changed the labor market? *The quarterly journal of economics*, **113**(4), 1169–1213.

Kim, J., & Marschke, G. 2005. Labor mobility of scientists, technological diffusion, and the firm's patenting decision. *Rand journal of economics 36, 298–317.*

Krakel, M., & Sliwka, D. 2009. Should you allow your employee to become your competitor? on noncompete agreements in employment contracts. *International economic review 50, 117–141.*

Lai, R., D'Amour, A., Doolin, D. M., Li, G.-C., Sun, Y., Torvik, V., & Yu, A. 2014. Disambiguation and co-authorship networks of the U.S. patent inventor database (1975-2010). *Research policy, elsevier*, **43**(6), 941–955.

Lemley, Mark A, & Sampat, Bhaven. 2012. Examiner characteristics and patent office outcomes. *Review of economics and statistics*, **94**(3), 817–827.

Machin, S., & Pekkarinen. 2008. Global sex differences in test score variability. *Science*, **322**(5906), 1331–1332.

Mankiw, N. Gregory. 2013. Defending the one percent. *Journal of economic perspectives*, **27**(3), 21–34.

Marx, M., Strumsky, D., & Fleming, L. 2009. Mobility, skills, and the michigan non-compete experiment. *Management science, 55 (6) 875-889.*

Matsuyama, Kiminori. 2002. The rise of mass consumption societies. *Journal of political economy*, **110**(5), 1035–1070.

McGranahan, Leslie, & Paulson, Anna. 2006. Constructing the chicago fed income based economic index–consumer price index: Inflation experiences by demographic group: 1983-2005. *Federal reserve bank of chicago working paper.*

McNally, S. 2013. Student awareness of costs and benefits of educational decisions: Effects of an information campaign. *Centre for economics of education discussion paper 139.*

Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, **71**(6), 1695–1725.

Melitz, Marc J, & Ottaviano, Gianmarco IP. 2008. Market size, trade, and productivity. *The review of economic studies*, **75**(1), 295–316.

Moen, J. 2005. Is mobility of technical personnel a source of r&d spillovers? *Journal of labor economics 23, 81–114.*

Moretti, Enrico. 2013. Real wage inequality. *American economic journal: Applied economics*, **5**(1), 65–103.

Nelson, R., & Phelps, E. 1966. Investment in humans, technological diffusion, and economic growth. *American economic review: 51 (2) 69-75.*

Pakes, A. 1986. Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica,54(4), 755-84.*

Pakes, A., & Nitzan, S. 1983. Optimum contracts for research personnel, research employment, and the establishment of rival enterprises. *Journal of labor economics, 1(4), 345-65.*

Pope, D., & Sydnor, J. 2010. Geographic variation in the gender differences in test scores. *Journal of economic perspectives, 24(2), 95-108.*

Romer, P. 2000. Should the government subsidize supply or demand in the market for scientists and engineers? *Nber chapters, in : Innovation policy and the economy, volume 1, 221-252.*

S., Samila, & Sorensen, O. 2011. Noncompete covenants: incentives to innovate or impediments to growth. *Management science (forthcoming).*

Sampat, Bhaven, & Williams, Heidi L. 2015. *How do patents affect follow-on innovation? evidence from the human genome.* Tech. rept. National Bureau of Economic Research.

Sato, Kazuo. 1976. The ideal log-change index number. *The review of economics and statistics*, 223–228.

Saxenian, A. 1994. Regional advantage: culture and competition in silicon valley and route 128. *Cambridge, massachusetts: Harvard university press.*

Schmookler, J. 1957. Inventors past and present. *Review of economics and statistics 39(3) 321-333.*

Schumpeter, Joseph A. 1942. *Socialism and democracy.*

Scotchmer, S. 2004. Innovation and incentives. *Cambridge: Ma free press.*

Shalem, R., & Trajtenberg, M. 2009. Software patents, inventors and mobility. *mimeo, tel aviv university.*

Song, Jae, Price, David J, Guvenen, Fatih, Bloom, Nicholas, & von Wachter, Till. 2015. *Firming up inequality.* Tech. rept. National Bureau of Economic Research.

Stern, S. 2004. Do scientists pay to be scientists? *Management science, 50(6) 835-853.*

Stoyanov, A., & Zubanov, N. 2012. Productivity spillovers across firms through worker mobility. *American economic journal: Applied economics, american economic association, 4: 168-98.*

Thursby, M., & Thursby, J. 2005. Gender patterns of research and licensing activity of science and engineering faculty. *Journal of technology transfer, 30(4), 343-353.*

Toivanen, O., & Vaananen, L. 2012. Returns to inventors. *Review of economics and statistics, 94(4), 1173-1190.*

Toivanen, O., & Vaananen, L. 2015. Education and invention. *Review of economics and statistics, forthcoming.*

Trajtenberg, M., Shiff, G., & Melamed, R. 2006. The "names game" : harnessing inventors' patent data for economic research. *Nber working paper no. 12479, cambridge, ma.*

Van Reenen, J. 1996. The creation and capture of economic rents: Wages and innovation in a panel of uk companies. *Quarterly journal of economics (1996) cxi, 443, 195-226.*

Vartia, Yrjö O. 1976. Ideal log-change index numbers. *Scandinavian journal of statistics*, 121–126.

Vernon, Raymond. 1966. International investment and international trade in the product cycle. *The quarterly journal of economics*, 190–207.

Waldinger, Fabian. 2010. Quality matters: The expulsion of professors and the consequences for phd student outcomes in nazi germany. *Journal of political economy*, **118**(4), 787–831.

Waldinger, Fabian. 2011. Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *The review of economic studies*, rdr029.

Wuchty, Stefan, Jones, Benjamin F, & Uzzi, Brian. 2007. The increasing dominance of teams in production of knowledge. *Science*, **316**(5827), 1036–1039.

Zhelobodko, Evgeny, Kokovin, Sergey, Parenti, Mathieu, & Thisse, Jacques-Francois. 2012. Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica*, **80**(6), 2765–2784.

Zucker, L., & Darby, M. 2014. Movement of star scientists and engineers and high-tech firm entry. *Nber working paper 12172.*

Zucker, L. G., Darby, M. R., & Brewer, M. B. 1998. Intellectual human capital and the birth of u.s. biotechnology enterprises. *American economic review, 88(1), 290-306.*

A. Appendix to Chapter 1

A.1. More on Theory

A. The Strong Equilibrium Bias in a Model Following Acemoglu (2002)

Consumers

Individual Demands

Preferences are different for the two groups:

$U_{poor} = a + u(x_L) = a + \frac{1}{1-\alpha_L} x_L^{1-\alpha_L}$

$U_{rich} = a + v(x_R) = a + \frac{1}{1-\alpha_R} x_R^{1-\alpha_R}$

where a is the numeraire. The FOC for the composite high-quality and low-quality goods are:

$u'(x_L) = x_R^{-\alpha_L} = p_L$

$v'(x_R) = x_R^{-\alpha_R} = p_R$

Nominal inequality is measured by w_{rich}/w_{poor}, while real inequality is measured by $\frac{w_{rich}}{p_H} / \frac{w_{poor}}{p_L}$.

Aggregate Demands

Normalize the total mass of consumers to 1, with share λ of rich types (if desired, we can introduce another scaling factor to study market size effect due to the total number of consumers). Aggregate demands are given by:

$D_L = (1-\lambda)p_L^{-\frac{1}{\alpha_L}}$

$D_H = \lambda p_H^{-\frac{1}{\alpha_H}}$

Producers

Final Producer

Final producer just uses capital and combines "varieties $x(v,i)$" to produce two types of goods, H or L (profits are thrown away). With $i = L/H$, we can write the problem as:

$max_{|x(v,i)| \in N(i)} \frac{p(i)}{1-\epsilon_i} \left(\int_0^{N(i)} x(v,i)^{1-\epsilon_i} dv \right) - \int_0^{N(i)} p^x(v,i)x(v,i)dv$

Note that the returns to scale are decreasing. The optimal choice is:

$x(v,i) = \left(\frac{p(i)}{p^x(v,i)} \right)^{\frac{1}{\epsilon_i}}$

We denote by $\sigma_i = \frac{1}{\epsilon_i}$ the elasticity of substitution between machines.

Intermediate Producers

Intermediate monopolist has a patent and chooses optimal price (we consider one period only here, but easy to extend since problem is separable). The cost of production of a machine is ψ_i units of the final good. The value of a patent (of a variety) for intermediate good in sector i (lasting one period) is:

$V(v,i) = (p^x(v,i) - \psi_i)x(v,t)$

where the optimal price chosen by the monopolist maximizes (taking demand as given):

$max_{p^x(v,i)}\ (p^x(v,i) - \psi_i) \left(\frac{p(i)}{p^x(v,i)}\right)^{\frac{1}{\epsilon_i}}$

Hence the optimal choice:

$p^x(v,i) = \frac{\psi_i}{1-\epsilon_i}$

The value function at the optimum is:

$V(v,i) = p(i)^{\frac{1}{\epsilon_i}} (\frac{1-\epsilon_i}{\psi_i})^{\frac{1-\epsilon_i}{\epsilon_i}} \epsilon_i$

Aggregate Supply

Total quantity supplied in equilibrium is given by

$:S_i^* = \frac{1}{1-\epsilon_i} \left(\int_0^{N(i)} x(v,i)^{1-\epsilon_i} dv\right) = \frac{1}{1-\epsilon} \left(\int_0^{N(i)} \left(\frac{p(i)}{p^x(v,i)}\right)^{\frac{1-\epsilon_i}{\epsilon_i}} dv\right) = \left(\frac{1}{1-\epsilon_i}\right)^{\frac{1}{\epsilon_i}} \left(\frac{p(i)}{\psi_i}\right)^{\frac{1-\epsilon_i}{\epsilon_i}} N(i)$

Solving for the equilibrium

With exogenous varieties

$p_H = \left[\left(\frac{\lambda}{N_H}\right) (1-\epsilon_H)^{\frac{1}{\epsilon_H}} (\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}\right]^{\frac{1}{\frac{1-\epsilon_H}{\epsilon_H}+\frac{1}{\alpha_H}}}$

$p_L = \left[\left(\frac{1-\lambda}{N_L}\right) (1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}\right]^{\frac{1}{\frac{1-\epsilon_L}{\epsilon_L}+\frac{1}{\alpha_L}}}$

Relative market size *increases* the relative price, because supply is fixed. So real inequality when prices are endogenous is "lower" than real inequality when prices are exogenous.

With endogenous varieties

Interior solution

At an interior solution (there is research on both kinds of goods), the no arbitrage condition between two types of inventions requires:

$$\eta V(v,H) = \eta V(v,L)$$

The equilibrium ratio of varieties and the equilibrium prices are given by:

$$\frac{(N_H)^{\frac{1}{\epsilon_H}}}{(N_L)^{\frac{1}{\epsilon_L}}} = \frac{(\lambda)^{\frac{1}{\epsilon_H}}}{(1-\lambda)^{\frac{1}{\epsilon_L}}} \left(\frac{\kappa_H}{\kappa_L}\right)^{\zeta} \frac{(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H^2}}}{(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L^2}}} \frac{(1-\epsilon_H)^{\left(\frac{1}{\epsilon_H}\right)^2}}{(1-\epsilon_L)^{\left(\frac{1}{\epsilon_L}\right)^2}}$$

$$p_H = \left(\frac{\psi_H}{1-\epsilon_H}\right)^{1-\epsilon_H} \left(\frac{1}{\epsilon_H}\right)^{\epsilon_H}$$

$$p_L = \left(\frac{\psi_L}{1-\epsilon_L}\right)^{1-\epsilon_L} \left(\frac{1}{\epsilon_L}\right)^{\epsilon_L}$$

$$V^* = 1$$

Corner solutions and the strong equilibrium bias

$$\begin{cases} V(v,H) > V(v,L) \\ N_H = N + \bar{N_H} \ \& \ N_L = \bar{N_L} \end{cases}$$

$$p_L = \left[\left(\frac{1-\lambda}{\bar{N_L}}\right)(1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}\right]^{\frac{1}{\frac{1-\epsilon_L}{\epsilon_L}+\frac{1}{\alpha_L}}}$$

Consider a relative demand shock (compared with previous periods, where the steady state relative demand is embodied in the steady state relative number of varieties).

Assume that all research is allocated to the high quality good:

$$V(v,H) > V(v,L) \ \ (Profitability\ Condition) \Longleftrightarrow \frac{\left(\frac{\lambda}{N+\bar{N_H}}\right)}{\left(\frac{1-\lambda}{\bar{N_L}}\right)} > \left[\frac{(\frac{1-\epsilon_L}{\psi_L})^{\frac{1-\epsilon_L}{\epsilon_L}} \epsilon_L}{(\frac{1-\epsilon_H}{\psi_H})^{\frac{1-\epsilon_H}{\epsilon_H}} \epsilon_H}\right]^{\zeta} \frac{(1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1}{\epsilon_H}} (\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}$$

There is "overshooting" of the relative price (strong equilibrium bias) if the new relative price is smaller than the old one:

$$\frac{p_H}{p_L} < \frac{\bar{p_H}}{\bar{p_L}} \ \ (Price\ Overshooting\ Condition) \Longleftrightarrow \frac{\frac{\lambda}{N+\bar{N_H}}}{\frac{1-\lambda}{\bar{N_L}}} < \left[\frac{(\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}{(\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}} \cdot \frac{(1-\epsilon_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1-\epsilon_H}{\epsilon_H}}} \cdot \frac{\epsilon_L}{\epsilon_H}\right]^{\zeta} \frac{(1-\epsilon_L)^{\frac{1}{\epsilon_L}} (\psi_L)^{\frac{1-\epsilon_L}{\epsilon_L}}}{(1-\epsilon_H)^{\frac{1}{\epsilon_H}} (\psi_H)^{\frac{1-\epsilon_H}{\epsilon_H}}}$$

which cannot be satisfied at the same time as the profitability condition.

B. The Strong Equilibrium Bias in a Model Following Melitz and Ottaviano (2008)

- Preferences for λ high-income consumers and $1-\lambda$ low-income consumers

$U_i = q^c_{0i} + \alpha \int_{\omega\in\Omega_i} q^c_{\omega i} d\omega - \frac{1}{2}\gamma \int_{\omega\in\Omega_i} (q^c_{\omega i})^2 d\omega - \frac{1}{2}\eta (\int_{\omega\in\Omega_i} q^c_{\omega i} d\omega)^2$

$\implies p_{\omega i} = \alpha - \gamma q^c_{\omega i} - \eta Q^c_i$; $Q_i = \int_{\omega\in\Omega_i} q^c_{\omega i} d\omega$

- To enter the differentiated sector, a firm must incur a sunk entry cost of f_E units of labor

  - Then the firm's unit labor requirement of cost c is drawn from a cumulative distribution function $G(c)$ with support on $[0, c_M]$

  - The zero-profit cost cutoff (c_{Di}) is a sufficient statistic that determines firm outcomes as a function of their cost draw:

$$\begin{aligned}
p_i(c) &= \frac{1}{2}(c_{Di} + c) \quad (prices)\\
\mu_i &= p_i(c) - c = \frac{1}{2}(c_{Di} - c) \quad (markups)\\
r_i(c) &= \frac{L_i}{4\gamma}[(c_{Di})^2 - c^2] \quad (revenues)\\
\pi_i(c) &= \frac{L_i}{4\gamma}[c_{Di} - c^2] \quad (profits)
\end{aligned}$$

- Under the assumption that productivity $\frac{1}{c}$ is Pareto distributed with lower bound $\frac{1}{c_M}$ and shape parameter k, the (closed economy) cost cutoff is given by $c_{Di} = (\frac{\gamma\phi}{L_i})^{\frac{1}{k+2}}$

  - So the cost cutoff falls (meaning the average productivity is higher) when varieties are closer substitutes (lower γ), when there is a better distribution of cost draws (lower c_M), when sunk costs fall (lower f_E) and in bigger markets (higher L_i). These comparative statics induce an increase in the "toughness of competition" in the form of a larger number of varieties consumed (higher N_i) and lower average prices (lower $\bar{p}_i$).

- This implies that the relative price decreases with (relative) market size
  - Intuition: firms are "locked in" and the long-run supply curve is downward sloping because of entry
  - Note that even in the short run relative prices will never mitigate the increase in inequality
    - Intuition: the marginal cost of production is constant: $p = \frac{\gamma}{L}q + \psi$
    - Can generate price effects mitigating increases in inequality by introducing specialized labor

A.2. More on Data

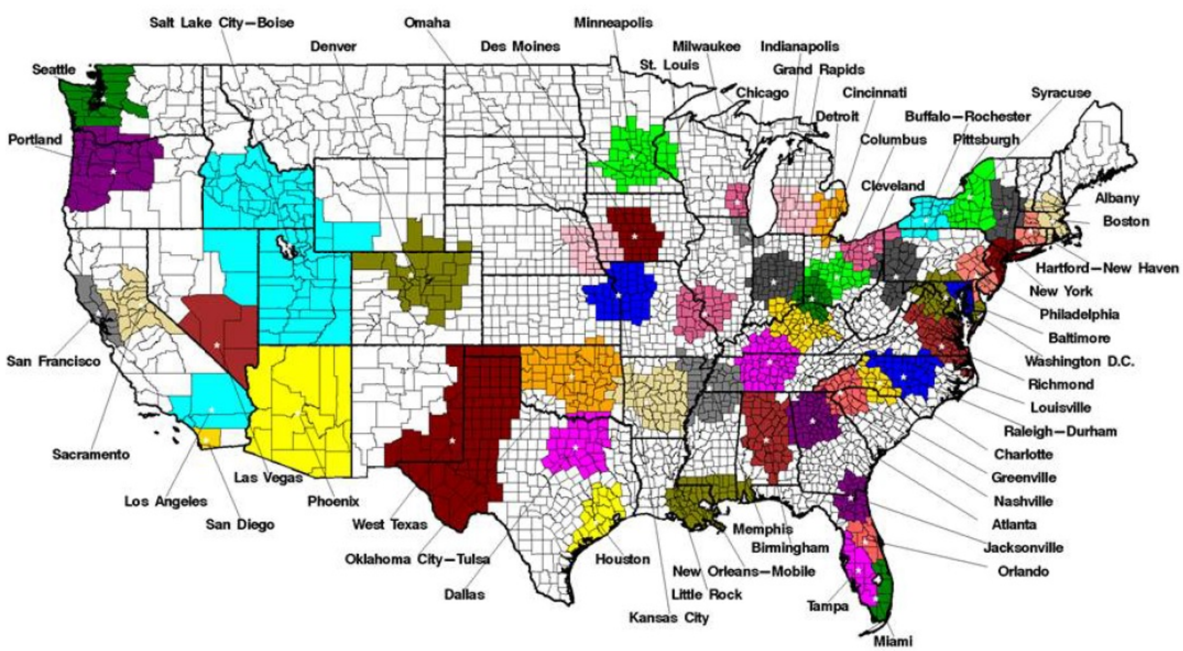
Description of Homescan Consumer Panel Data: I primarily rely on the Home Scanner Database collected by AC Nielsen and made available through the Kilts Center at The University of Chicago Booth School of Business. AC Nielsen collects these data using hand-held scanner devices that households use at home after their shopping in order to scan each individual transaction they have made. Faber and Fally (2015) report that on average each semester covers \$105 million worth of retail sales across 58,000 individual, across more than 500,000 barcodes belonging to 180,000 brands.

Description of Retail Scanner Data: The Retail Scanner Data consist of weekly price and quantity information for more than one hundred retail chains across all US markets between January 2006 and December 2013. The database includes about 45,000 individual stores. The stores in the database vary in terms of the channel they represent: e.g. food, drug, mass merchandising, liquor, or convenience stores. Faber and Fally (2015) report that on average each semester covers \$110 billion worth of retail sales across 25,000 individual stores, across more than 700,000 barcodes belonging to 170,000 brands.

The strength of the home scanner database is the detailed level of budget share information that it provides alongside household characteristics. Its relative weakness in the comparison to the store-level retail scanner data is that the home scanner samples households and, therefore, has higher sampling error at the product level. Relative to the home scanner data, the store-level retail scanner data records more than one thousand times the retail sales in each semester. I primarily rely on the home scanner data in the paper, but I present robustness checks based on the retail scanner data.

Local Markets: Both the home scanner and retail scanner data can be disaggregated into 76 local markets, which are shown on the map below.

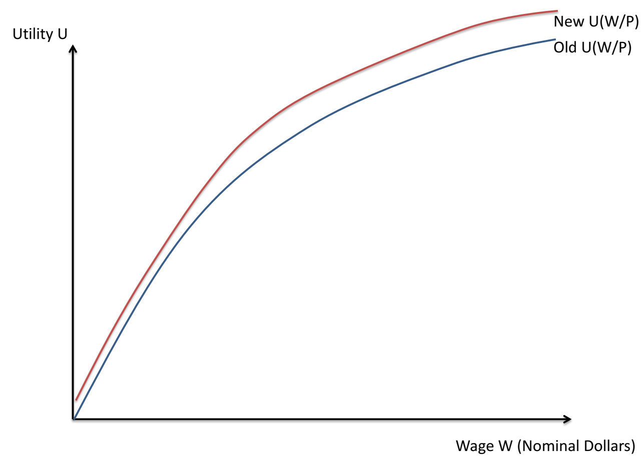
Figure A.1.: Map of the 76 Local Markets Tracked in the Nielsen Datasets



A.3. Estimation of Quality-Adjusted Inflation and Further Robustnes Checks

Nominal and Real Inequality

Figure A.2.: The Mapping Between Nominal Income and Utility



Estimation Equations

Estimating the elasticities: given the formula reported in the main text, we only need to estimate the group-specific and module-specific elasticities. We do this by first modeling the supply and demand conditions for each good within a module.

The demand equation comes from the following transformation, which exploits the panel nature of the data:

$$
\begin{aligned}
ln(s_{umgt}) - ln(s_{umg(t-1)}) &= \Delta ln(s_{umgt}) \\
&= (1-\sigma_m)\left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + ln(P_{mgt}) - ln(P_{mg(t-1)}) \\
&= (1-\sigma_m)\left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + \lambda_{mt}
\end{aligned}
$$

where the second line uses (1) and the fact that quality/taste is assumed to be constant over time. The fixed effect corresponds to the change in the price index of the module. In practice, there will be an estimation error, which for instance could come from yearly change in taste (which would affect the$d$ parameters). So we can write the demand curve as:

$$\Delta ln(s_{umgt}) = (1-\sigma_m)\Delta ln(p_{umgt}) + \lambda_{mt} + \epsilon_{umgt}$$

Then, we assume an isoelastic supply curve (with $\alpha > 0$ assumed to be the same for all UPCs within a module):

$$ln(c_{umgt}) = \alpha ln(p_{umgt}) + \chi_{mg}$$

$$ln(c_{umgt}) - ln(E_{mgt}) = \alpha ln(p_{umgt}) - ln(E_{mgt}) + \chi_{mg}$$

$$ln(s_{umgt}) = \alpha ln(p_{umgt}) - ln(E_{mgt}) + \chi_{mg}$$

Differencing over time:

$$
\begin{aligned}
ln(s_{umgt}) - ln(s_{umg(t-1)}) &= \Delta ln(s_{umgt}) \\
&= \alpha\left[ln(p_{umgt}) - ln(p_{umg(t-1)})\right] + ln(E_{mgt}) - ln(E_{mg(t-1)})
\end{aligned}
$$

so

$$
\begin{aligned}
\Delta ln(p_{umgt}) &= \frac{1}{\alpha}\Delta ln(s_{umgt}) - \frac{1}{\alpha}\Delta ln(E_{mgt}) \\
&= \frac{1}{\alpha}\Delta ln(s_{umgt}) + \psi_{mgt}
\end{aligned}
$$

The fixed effect corresponds to the change in total expenditures in the module (which is observed). In practice there will be estimation error, e.g. due to assembly line shocks, so we write:

$\Delta ln(p_{umgt}) = \frac{1}{\alpha}\Delta ln(s_{umgt}) + \psi_{mgt} + \delta_{umgt}$

We now want to eliminate the fixed effects in the demand and supply equations. We take a difference relative to the UPC k with the largest market share:

$\Delta^k ln(s_{umgt}) = (1-\sigma_m)\Delta^k ln(p_{umgt}) + \epsilon^k_{umgt}$

$\Delta^k ln(p_{umgt}) = \frac{1}{\alpha}\Delta^k ln(s_{umgt}) + \delta^k_{umgt}$

with $\Delta^k X = \Delta X_{umgt} - \Delta X_{kmgt}$, $\epsilon^k_{umgt} = \epsilon_{umgt} - \epsilon_{kmgt}$ and $\delta^k_{umgt} = \delta_{umgt} - \delta_{kmgt}$.

Now we can set up the moment condition, based on the assumption that the upc-specific demand and supply shocks are uncorrelated over time, i.e $E_t[\epsilon^k_{umgt}\delta^k_{umgt}] = 0$.

$v_{umgt} = \epsilon^k_{umgt} \times \delta^k_{umgt}$

$G(\beta_m) = E_t(v_{umgt}(\beta_m)) = 0 \;\; \forall u, m \; and \; g$

This can be written as:

$$
\begin{aligned}
v_{umgt}(\beta_m) &= \epsilon^k_{umgt} \times \delta^k_{umgt} \\
&= \left(\Delta^k ln(s_{umgt}) - (1-\sigma_m)\Delta^k ln(p_{umgt})\right) \times \left(\Delta^k ln(p_{umgt}) - \frac{1}{\alpha}\Delta^k ln(s_{umgt})\right) \\
&= \Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt}) - (1-\sigma_m)\left(\Delta^k ln(p_{umgt})\right)^2 - \frac{1}{\alpha}\left(\Delta^k ln(s_{umgt})\right)^2 \\
&\quad + \frac{(1-\sigma_m)}{\alpha}\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt}) \\
&= (\sigma_m - 1)\left(\Delta^k ln(p_{umgt})\right)^2 - \frac{1}{\alpha}\left(\Delta^k ln(s_{umgt})\right)^2 + \frac{\alpha + (1-\sigma_m)}{\alpha}\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})
\end{aligned}
$$

The moment condition $E_t[v_{umgt}(\beta_m)] = 0$ means:

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{1}{\alpha(\sigma_m-1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{\alpha+(1-\sigma_m)}{\alpha(\sigma_m-1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right] \quad \forall u, m \; and \; g$$

Rewriting $\alpha = \frac{1+\omega_m}{\omega_m}$ to match the notation in Broda and Weinstein (2006), this yields:

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{\omega_m}{(1+\omega_m)(\sigma_m-1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{(1+\omega_m)+(1-\sigma_m)\omega_m}{(1+\omega_m)(\sigma_m-1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right]$$

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \frac{\omega_m}{(1+\omega_m)(\sigma_m-1)} E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \frac{1-\omega_m(\sigma_m-2)}{(1+\omega_m)(\sigma_m-1)} E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right]$$

$$E_t\left[\left(\Delta^k ln(p_{umgt})\right)^2\right] = \theta_1 E_t\left[\left(\Delta^k ln(s_{umgt})\right)^2\right] - \theta_2 E_t\left[\Delta^k ln(s_{umgt})\Delta^k ln(p_{umgt})\right]$$

We then estimate the parameters ω_m and σ_m, under the restriction that $\omega_m > 0$ and $\sigma_m > 1$. To do this, we first just estimate θ_1 and θ_2 by weighted least squares, as in Feenstra (1994). Then we go back to the primitive parameters. If this produces imaginary estimates or estimates of the wrong sign, we perform grid search for the objective function for values of $\sigma_g \in [1.05, 131.5]$ at intervals that are 5 percent apart.

Average Inflation Rate of Various Income Groups According to Various Price Indices

Table A.1.: Average Annual Inflation Rates Across Three Income Groups

Panel A: Full Sample (Percentage Points)

	Income < \$30k		Income ∈ [\$30k-\$100k]		Income > \$100k	
	Arith. Avg.	Geom. Avg.	Arith. Avg.	Geom. Avg.	Arith. Avg.	Geom. Avg.
Geometric Laspeyres	1.212	1.204	0.912	0.951	0.561	0.639
Truncated Geometric Laspeyres	1.544	1.536	1.137	1.157	0.862	0.909
Paasche	1.580	1.571	0.985	1.010	0.965	0.979
Truncated Paasche	1.719	1.710	1.182	1.194	1.117	1.126
Tornqvist	1.938	1.929	1.426	1.418	1.296	1.290
Fisher	1.983	1.974	1.425	1.418	1.327	1.320
Marshall-Edgeworth	1.992	1.984	1.440	1.433	1.330	1.323
CES Ideal	2.041	2.032	1.529	1.522	1.387	1.380
Truncated CES Ideal	2.063	2.054	1.541	1.534	1.413	1.406
Walsh	2.076	2.067	1.571	1.563	1.423	1.416
Truncated Laspeyres	2.257	2.502	1.724	1.910	1.554	1.721
Laspeyres	2.387	2.379	1.867	1.860	1.689	1.682
Truncated Geometric Paasche	2.433	2.424	1.742	1.734	1.822	1.815
Geometric Paasche	2.669	2.660	1.942	1.934	2.037	2.031

Panel B: All Years but Great Recession (Percentage Points, Arithmetic Average)

	Income < \$30k	Income ∈ [\$30k-\$100k]	Income > \$100k
Geometric Laspeyres	0.870	0.642	0.318
Truncated Geometric Laspeyres	1.179	0.876	0.627
Paasche	1.246	0.732	0.768
Truncated Paasche	1.380	0.928	0.919
Tornqvist	1.586	1.144	1.085
Fisher	1.625	1.161	1.111
Marshall-Edgeworth	1.633	1.176	1.116
CES Ideal	1.674	1.254	1.169
Truncated CES Ideal	1.695	1.268	1.192
Walsh	1.707	1.297	1.204
Truncated Laspeyres	1.891	1.448	1.316
Laspeyres	2.006	1.592	1.455
Truncated Geometric Paasche	2.071	1.467	1.623
Geometric Paasche	2.308	1.648	1.858

Table A.2.: Average Annual Inflation Rates Across Four Income Groups

Panel A: Full Sample (Percentage Points)

	Income < \$25k	Income ∈ [\$25k-\$50k]	Income ∈ [\$50k-\$100k]	Income > \$100k
Geometric Laspeyres	1.236	1.029	0.785	0.561
Truncated Geometric Laspeyres	1.561	1.293	1.025	0.862
Paasche	1.647	1.249	0.962	0.965
Truncated Paasche	1.766	1.414	1.132	1.117
Tornqvist	2.000	1.668	1.365	1.296
Fisher	2.045	1.687	1.377	1.327
Marshall-Edgeworth	2.052	1.698	1.396	1.330
CES Ideal	2.086	1.763	1.462	1.387
Truncated CES Ideal	2.106	1.778	1.474	1.413
Walsh	2.116	1.800	1.501	1.423
Truncated Laspeyres	2.293	1.984	1.657	1.554
Laspeyres	2.445	2.126	1.795	1.689
Truncated Geometric Paasche	2.527	2.090	1.738	1.822
Geometric Paasche	2.769	2.311	1.949	2.037

Panel B: All Years but Great Recession (Percentage Points, Arithmetic Average)

	Income < \$25k	Income ∈ [\$25k-\$50k]	Income ∈ [\$50k-\$100k]	Income > \$100k
Geometric Laspeyres	0.843	0.729	0.529	0.318
Truncated Geometric Laspeyres	1.164	1.007	0.769	0.627
Paasche	1.289	0.964	0.723	0.768
Truncated Paasche	1.405	1.148	0.885	0.919
Tornqvist	1.613	1.374	1.097	1.085
Fisher	1.660	1.392	1.127	1.111
Marshall-Edgeworth	1.666	1.403	1.148	1.116
CES Ideal	1.692	1.469	1.201	1.169
Truncated CES Ideal	1.713	1.489	1.211	1.192
Walsh	1.722	1.504	1.241	1.204
Truncated Laspeyres	1.895	1.683	1.394	1.316
Laspeyres	2.033	1.823	1.533	1.455
Truncated Geometric Paasche	2.143	1.805	1.474	1.623
Geometric Paasche	2.388	2.024	1.669	1.858

Table A.3.: Average Annual Inflation Rates across the Income Distribution at UPC*Geography Level

	Income < \$30k	Income $\in$ [\$30k-\$100k]	Income > \$100k
Paasche	2.065	1.401	1.341
CES Ideal	2.434	1.902	1.722
Laspeyres	2.789	2.365	2.08

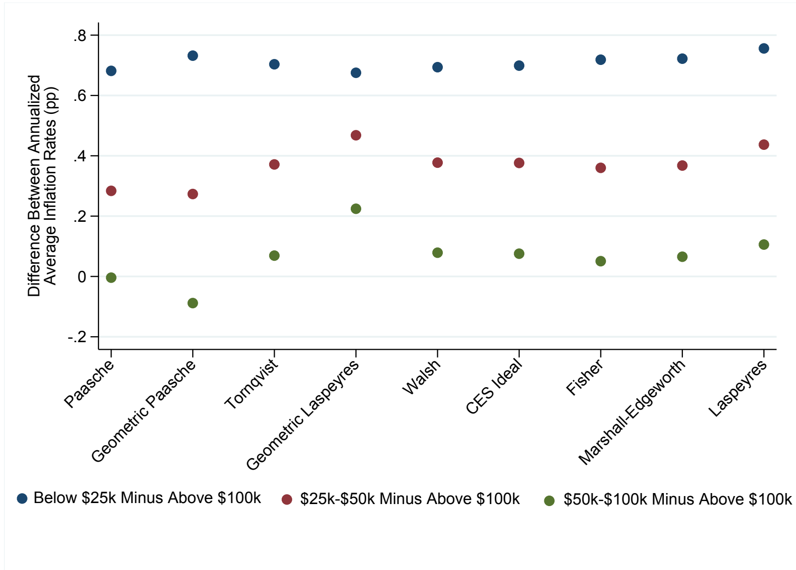
Table A.4.: Average Annual Inflation Rates across the Income Distribution at UPC*Store Level

	Income < \$30k	Income $\in$ [\$30k-\$100k]	Income > \$100k
Paasche	2.239	2.002	1.692
CES Ideal	2.471	2.248	1.901
Laspeyres	2.710	2.471	2.072

Table A.5.: Average Annual Inflation Rates across the Income Distribution at Quarterly Level

	Income < \$30k	Income $\in$ [\$30k-\$100k]	Income > \$100k
Paasche	-1.161	-2.268	-2.124
CES Ideal	1.911	1.107	1.066
Laspeyres	5.429	5.042	4.956

Figure A.3.: Inflation Difference Between Various Income Groups For Various Price Indices (Fixed Basket)



Is Differential Inflation (Fixed Basket) Across the Income Distribution Driven by a Selection Effect?

Table A.6.: Products that are about to exit have a lower inflation rate

Subsample	Laspeyres Inflation Rate	Median Laspeyres Inflation Rate
Continued	2.03%	2.06%
About to Exit	-1.33%	-0.52%
Justed Entered	0.03%	1.3%

Table A.7.: Products that are about to exit have a higher price level

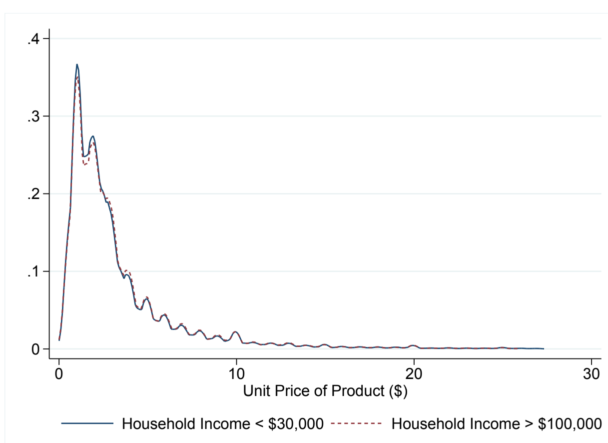
Subsample	Average Price Level	Median Price Level
Continued	3.67	2.75
About to Exit	3.95	2.68
Justed Entered	4.91	3.05

Table A.8.: Share of spending on new and discontinued products across the income distribution

	Share of Spending on Products...	
Household Income	About to Exit	Just Entered
$> \$100,000$	3.04%	10.94%
$\$30,000 - \$100,000$	2.71%	10.01%
$< \$30,000$	2.59%	9.26%

Differences in Prices Paid for Same Products for Rich and Poor

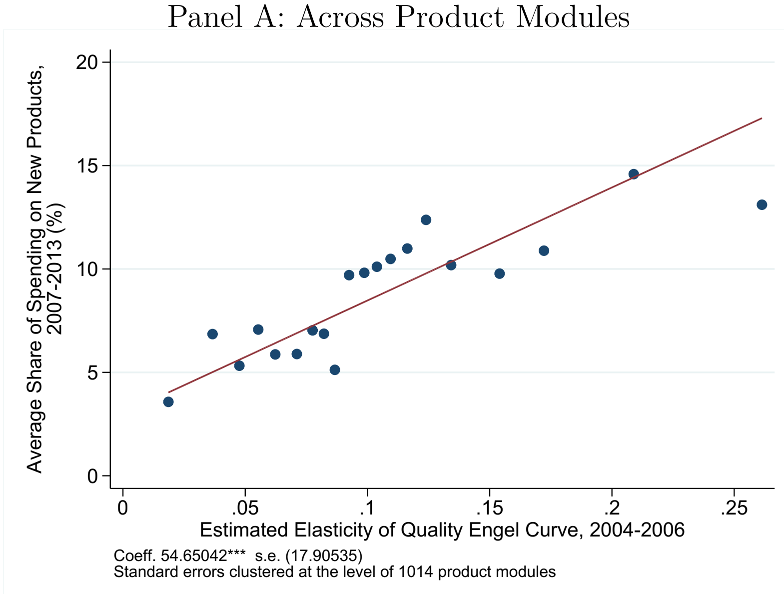
Figure A.4.: The Distribution of Average Unit Prices Paid is the Same Across the Income Distribution (Reweighting by Spending Shares)



Further Robustness Checks on Increase in Product Variety Across the Income Distribution

Figure A.5.: The Positive Relationship Between Share of Spending on New Products and Mean Consumer Income

Panel A: Across Product Modules



Panel B: Across Product Modules with Product Group Fixed Effects

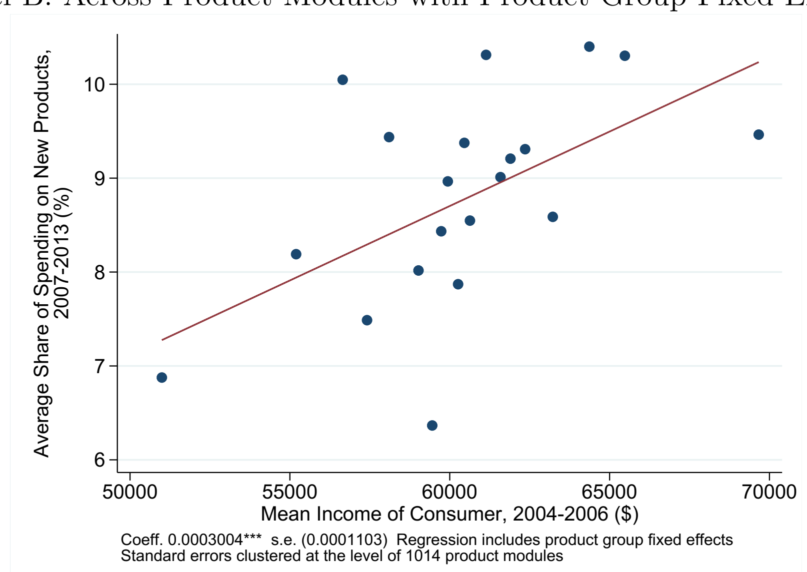


Figure A.6.: The Relationship Between Share of Spending on New Products and Mean Income Depends on the Quality Engel Curves Elasticity

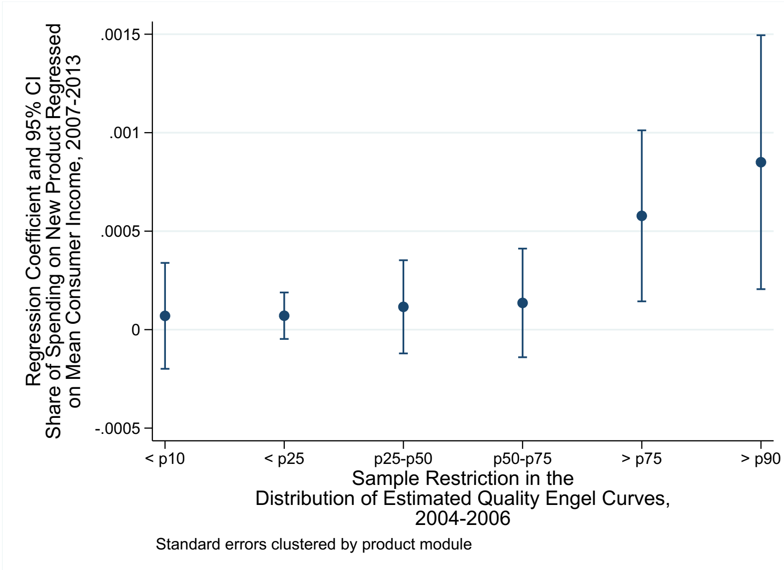
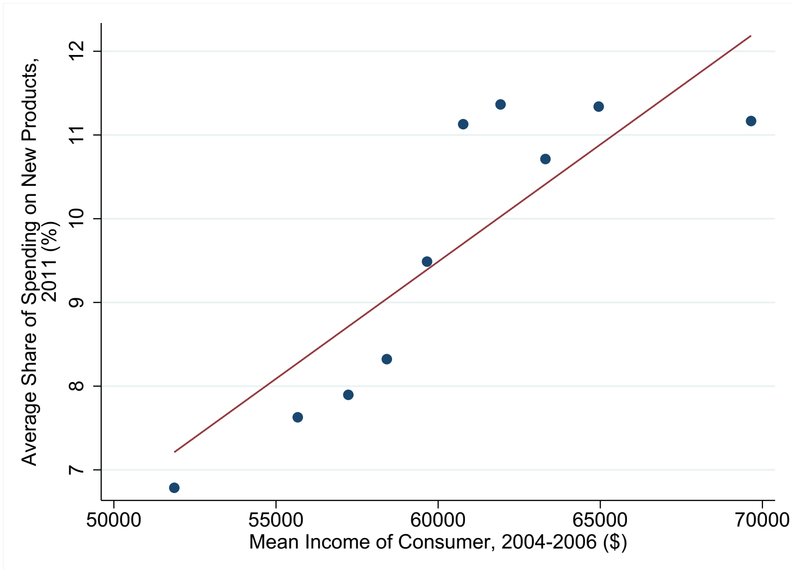


Figure A.7.: The Relationship Between Share of Spending on New Products and Representative Consumer Mean Income, Controlling for Household Fixed Effects

Panel A: Relationship in the Full Sample



Panel B: Relationship across the Distribution of Quality Engel Curve Elasticities

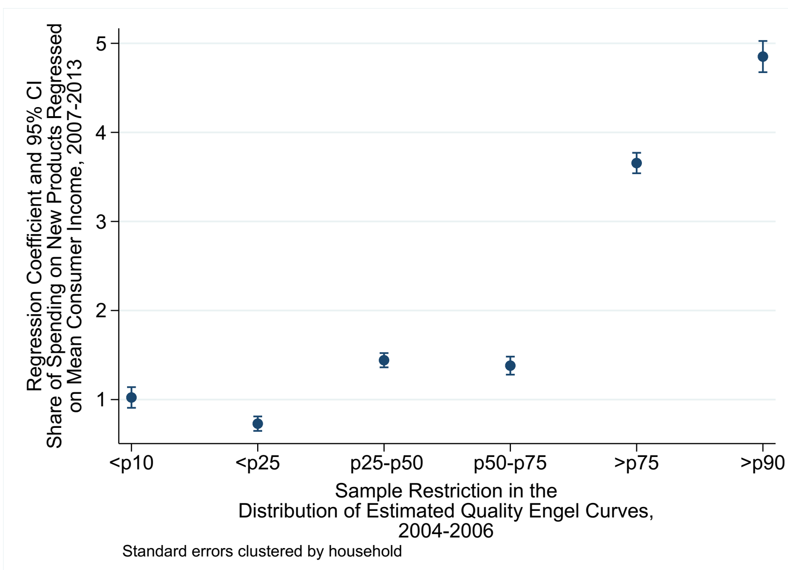
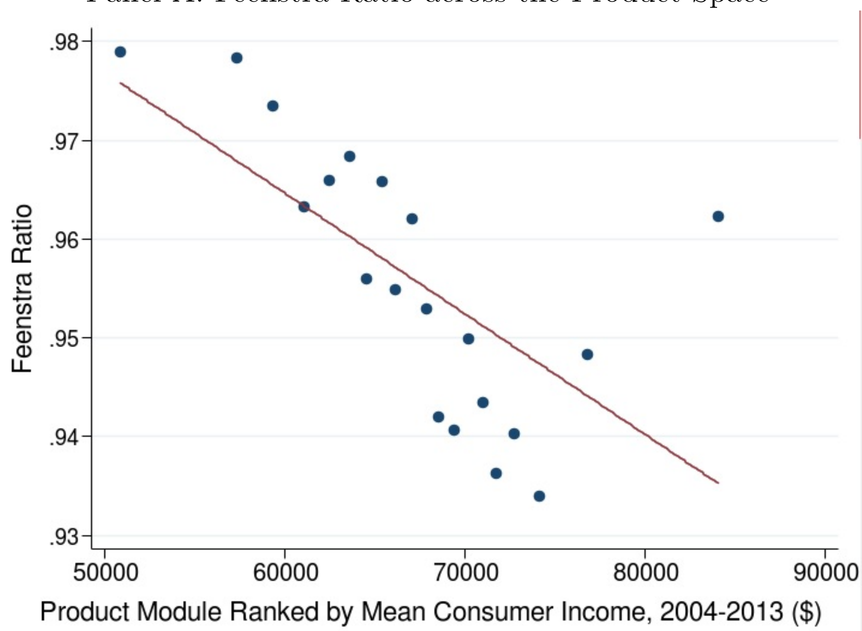


Figure A.8.: Robustness Checks on Increase in Product Variety across the Income Distribution

Panel A: Feenstra Ratio across the Product Space



Panel B: Annual Growth in Total UPC Count across the Product Space

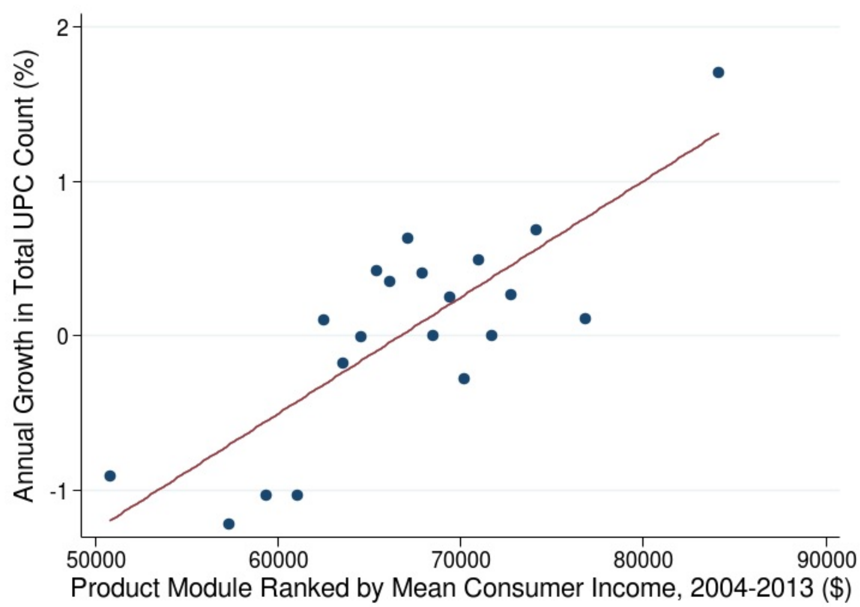


Table A.9.: Decomposing the Difference in Sahares of Spending between High- and Low-Income Households

Aggregation Level (Broad to Narrow)	Decomposition	Difference (% of actual)
Department	Between	**1.8**
Product Group	Between	**29.0**
Product Module	Between	**39.2**

Quality-Adjusted Inflation

Figure A.9.: Elasticities of Substitution Differ Across the Income Distribution

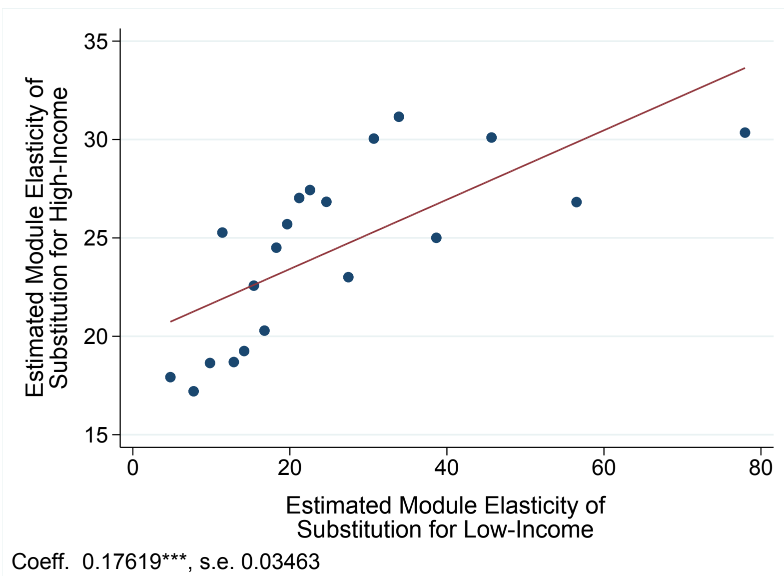


Figure A.10.: Quality-Adjusted Inflation Across the Income Distribution

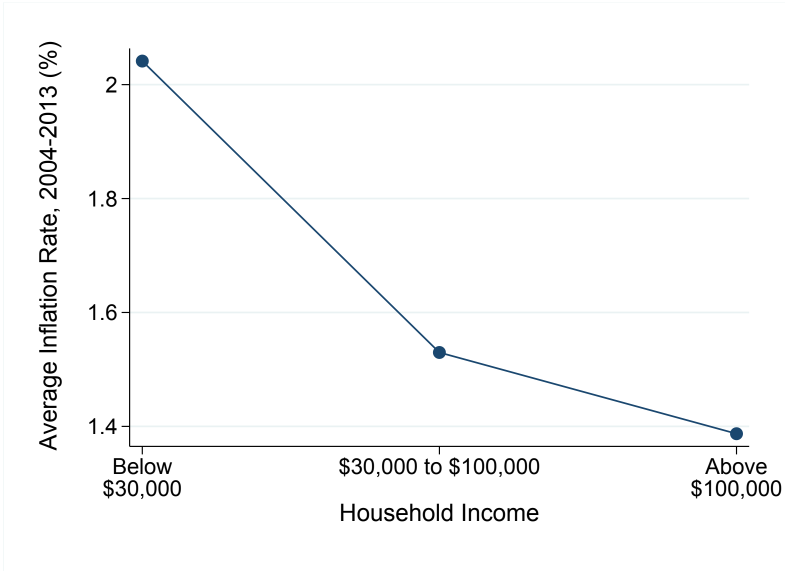


Table A.10.: Isolating the Contribution of Stores and Local Market to the Overall Inflation Difference

Price Change	Local Market Shares	Store Shares	UPC Shares	Inflation Difference (% of Benchmark)
Counterfactual	Actual	Actual	Actual	**100**
Counterfactual	Counterfactual	Actual	Counterfactual	**43.2**
Counterfactual	Actual	Counterfactual	Counterfactual	**3.1**

Notes: This table is based on a reweighting methodology that expresses the spending share of a product as a combination of the spending shares of the local market, of the store within that market, and of the barcode within that store. The counterfactal share explained by store effects is an upper bound because for some stores, only high-income or only low-income consumers are observed. $P_L \equiv \sum_{i=1}^{n} \frac{p_i^t}{p_i^0} s_{local\ marlet} \cdot s_{store} \cdot s_{upc}$

A.4. Robustness Checks on the Relationship between Changes in Market Size and Quality-Adjusted Inflation

Additional Motivating Evidence

Figure A.11.: The Share of Rich Households is Primarily Correlated with High Price Quantiles of New Products

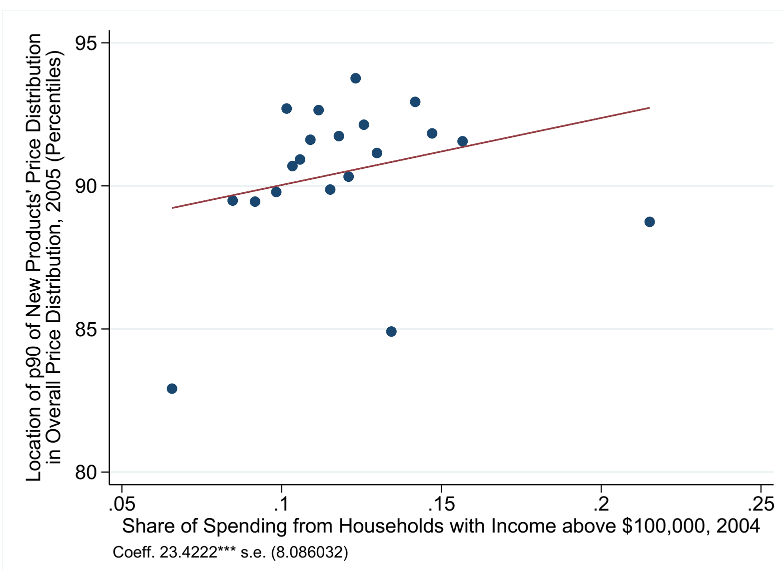
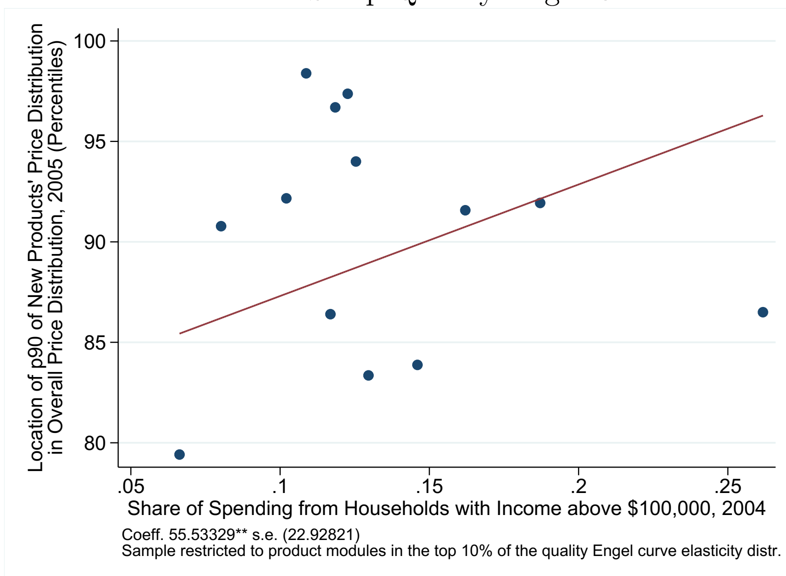
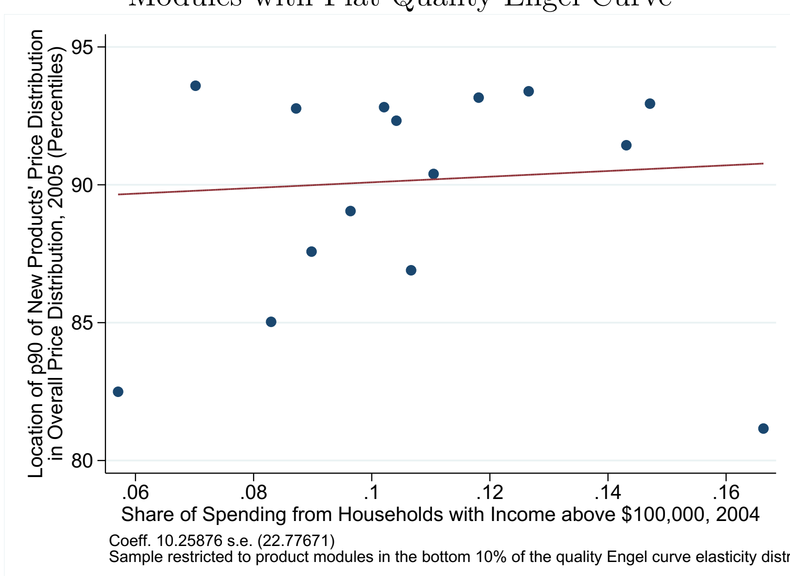


Figure A.12.: The Share of Rich Households is Correlated with High Price Quantiles of New Products only in Product Modules with Steep Quality Engel Curves

Panel A: Location of p90 of New Products in Overall Price Distribution for Product Modules with Steep Quality Engel Curves



Panel B:Location of p90 of New Products in Overall Price Distribution for Product Modules with Flat Quality Engel Curve



Additional Results on National Research Design

Figure A.13.: Predicted and Actual Increase in Market Size

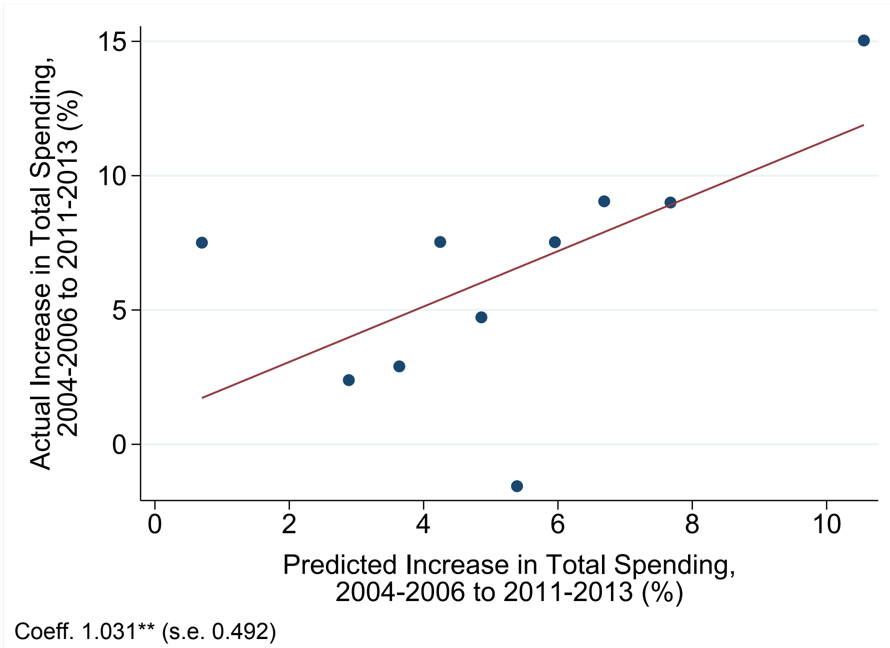


Table A.11.: Further Robustness Checks on Causal Effects of Changes in Market Size

Panel A: Controlling for 2004-2006 Age and Income Distributions and Price Decile Fixed Effects

	Share of Spending on New Products (pp)			Overlapping Goods Inflation Rate (pp)		
Predicted Increase in Spending (%)	0.527***	0.380***	0.419***	-0.159***	-0.137***	-0.125***
	(0.072)	(0.144)	(0.179)	(0.022)	(0.037)	(0.038)
Age Distribution Controls	Yes	Yes	Yes	Yes	Yes	Yes
Income Distribution Controls	No	Yes	Yes	No	Yes	Yes
Price Decile Fixed Effects	No	No	Yes	No	No	Yes
Product Module Fixed Effects	No	No	No	No	No	No
Spending Weights	Yes	Yes	Yes	Yes	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	9,089	9,089	9,089	9,089	9,089	9,089
Number of Clusters	1,006	1,006	1,006	1,006	1,006	1,006

Standard errors clustered by product modules.

Panel B: Robustness to Other Weights

	Share of Spending on New Products (pp)	Overlapping Goods Inflation Rate (pp)
Predicted Increase in Spending (%)	0.296***	-0.086***
	(0.0585)	(0.0155)
Product Module Fixed Effects	Yes	Yes
Spending Weights	Yes	Yes
Sample Restricted to Positive Spending Growth	Yes	Yes
Number of Observations	8,545	8,545
Number of Clusters	1,000	1,000

Sample restricted to product modules below 95th percentile of total spending.
Standard errors clustered by product modules.

Results from Geography Research Design

As a robustness check, I use time variation in the age and income distribution of households in 76 local markets tracked by Nielsen within the US (see Appendix B for details). A local market is a county group defined by Nielsen, which I match to local covariates from the American Community Survey. For each product module in each local market, I predict change in market size based on local change in age and income distributions. Some cities like San Francisco have experienced an increasing share of high-income and young households, while in other cities like New Orleans have become poorer, with a decline in overall population. I then compare the change in fixed basket inflation across product module$\times$ local market cells with increasing or decreasing predicted market size. To control for supply factors, I include fixed effects: local market fixed effects control for local scale effects, while product group fixed effect control for national trends in inflation. In this setting, the identification assumption is that, conditional on the fixed effects, the direct effect of local changes in the age and income distributions on the equilibrium is only through demand.

I use 18 covariates X_{it} (all expressed in logs): total number of households, total population, total female population, total male population, total number of households in age $\times$ income groups (considering four age groups - below 25, 25 to 44, 45 to 64, above 65 -, and dividing each group into 3 income groups - below \$30k, \$30k to \$100k, above \$100k), median household income and mean household income.

Formally, I consider two periods, 2004-2006 and 2011-2013, and I predict (log) local total expenditures Q_{MIT} with local market covariates and fixed effects:

$$Q_{MIT} = \beta^M \cdot 1_M \cdot X_{IT} + \gamma_{IT} + \delta_{GT} + \epsilon_{IMT}$$

where M denotes the product module, G the product group, I the local market, and T the period. Note that the β^M coefficients are allowed to freely vary across modules (i.e. some modules will be very responsive to the number of low-income households, others more responsive to the number of high-income and old households, etc.). I estimate this specification in $2004-2006$ and predict market size out-of-sample in $2011-2013$ (the R^2 is very high: see Appendix D).

The predictor of residual market size growth between two periods is therefore $\beta^m \cdot 1_M \cdot (X_{IT_2} - X_{IT_1})$. Finally, I run specifications of the form:

$$Y_{MI} = \alpha\left[\beta^m \cdot 1_M(X_{IT_2} - X_{IT_1})\right] + \widetilde{\gamma_I} + \widetilde{\delta_G} + \widetilde{\epsilon_{IM}}$$

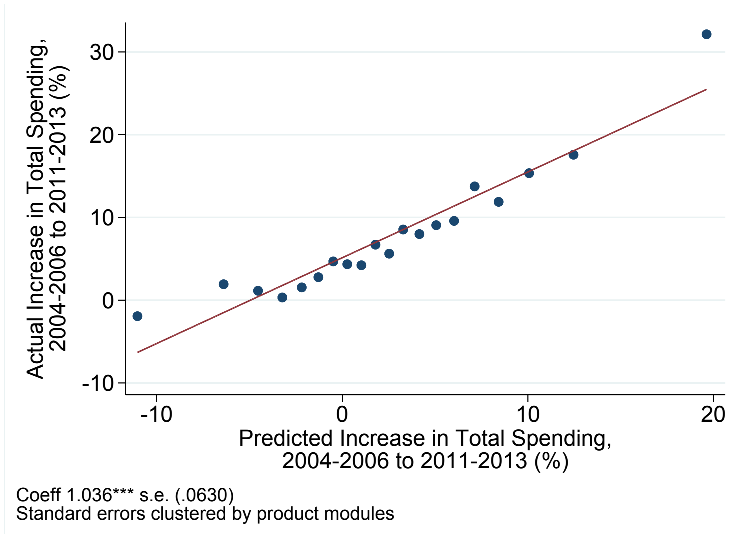
Figure A.14 and the table below show that the relationship is very stable and very similar to the results found from the variation at the national level.

Table A.12.: Causal Effects of Changes in Market Size (Local Level)

	Difference in Fixed Basket Inflation Rate (pp)			
Predicted Increase in Spending (%)	-0.1471***	-0.1276***	-0.1271***	-0.1276***
	(0.0162)	(0.0172)	(0.0188)	(0.0259)
F.E.	Department	Product Group	Product Group	Product Group
Weights	Yes	Yes	No	Yes
Cluster	Local Market	Local Market	Local Market	Product Module

Figure A.14.: Causal Effects of Changes in Market Size (Local Level)

A. Predicted and Actual Increase in Market Size



B. Higher Market Size Leads to Lower Inflation (Fixed Basket)

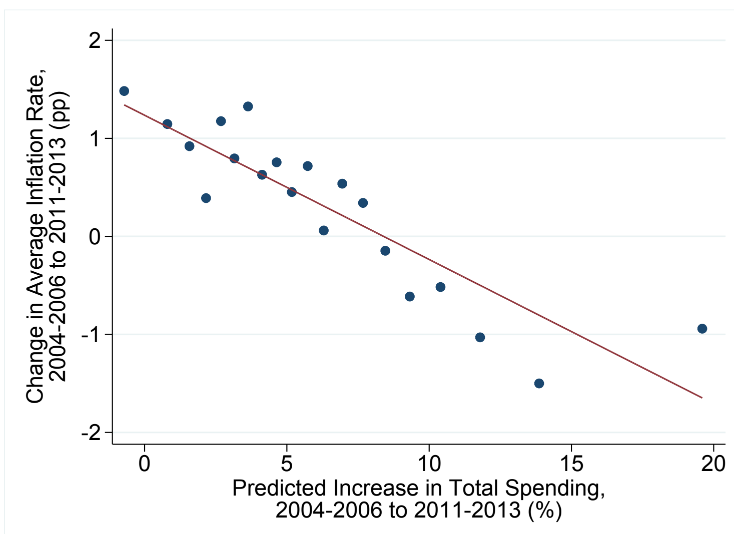
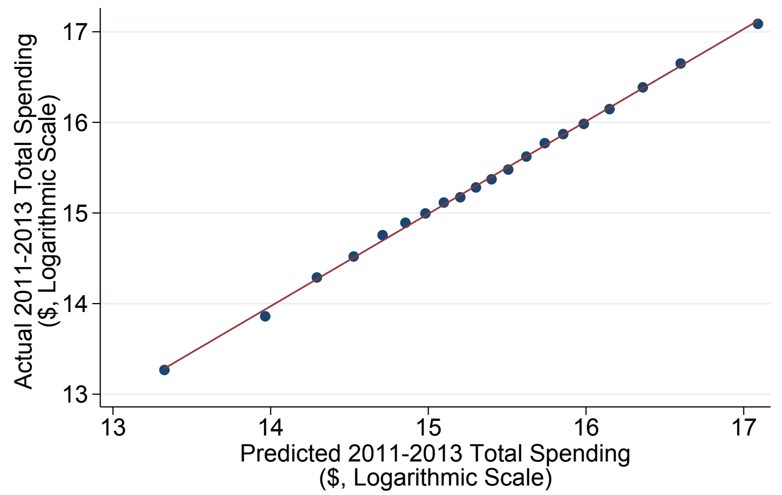
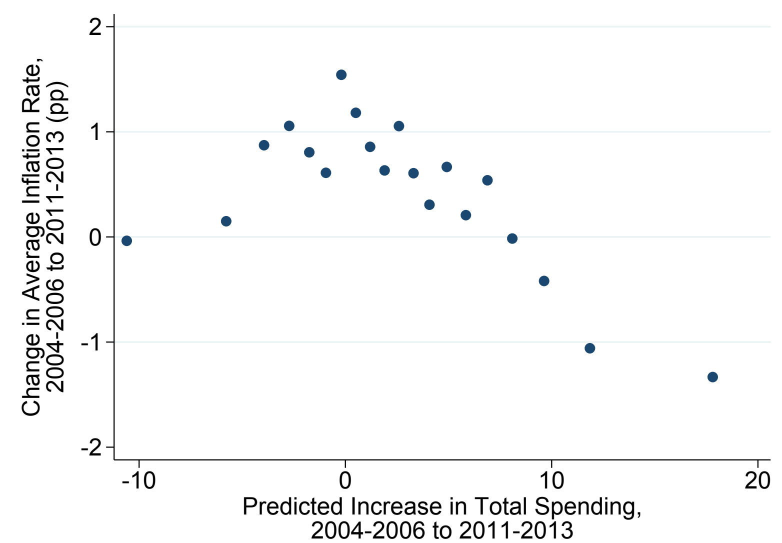


Figure A.15.: Geography Design: Predicting Market Size



Note: Out-of-sample prediction using the prediction function estimated on 2004-2006 data and partialling out scantrack market and product group fixed effects

Figure A.16.: Geography Design: Predicted Market Size Growth and Fixed-Basket Inflation



Additional Evidence on Mechanism

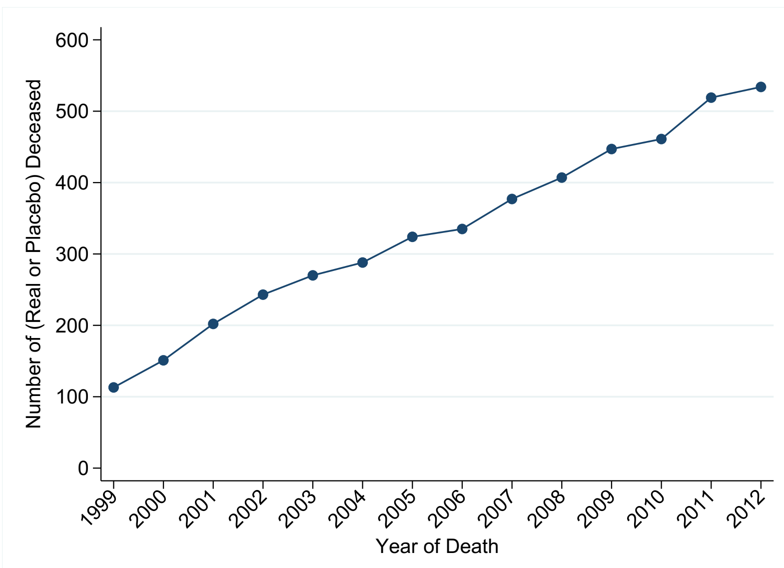
Figure A.17.: Do Product Innovations Follow Market Size or Change in Market Size?

	Share of Spending on New Products	
Lagged Change in Market Size	3.107***	1.901**
	(1.139)	(0.926)
Lagged Market Size	1.399	0.577
	(1.439)	(1.269)
Product Group Fixed Effects	No	Yes
Weights	Yes	Yes

B. Appendix to Chapter 2

B.1. Additional Summary Satistics on Matched Inventors

Figure B.1.: Number of Deceased Inventors Per Year



Notes: This figure shows, in each year between 1999 and 2012, the number of inventors who passed away before the age of 60 and who had at least one co-inventor. The reason why the number of deceased inventors per year is increasing over time is that, for a deceased inventor to become part of our analysis, they need to have applied for at least one co-invented patent between 1996 and the year of their death (otherwise they have no associated survivor inventor). More and more inventors have applied for co-invented patents as we get closer to 2012, the end of our sample, therefore the number of deceased inventors per year is increasing over time.

Table B.1.: Balance in Technology Classes For Survivor Co-inventors

	Share of Patents at Co-inventor Death	
Technology Class	Real	Placebo
1. Chemical	14.37	14.82
2. Computers & Communications	28.60	27.49
3. Drugs & Medical	15.05	14.50
4. Electrical & Electronic	14.99	15.39
5. Mechanical	13.20	13.82
6. Others	13.58	13.61

Notes: This table shows the breakdown by technology class of all patents the real and placebo survivor inventors had invented at the time of their co-inventor death. The table shows very good balance across the two groups, although we did not use this information for the match described in Section II.B.

Table B.2.: Additional Balance Tests

Variable	Sample	Mean	SD	10pc	25pc	50pc	75pc	90pc
Number of Co-inventors	Real Survivors	9.726	10.85	2	3	6	12	21
	Placebo Survivors	9.583	10.61	2	3	6	12	21
	Real Deceased	3.002	3.873	1	1	2	5	10
	Placebo Deceased	2.83199	3.423	1	1	2	5	9
Firm Size	Real Survivors	35,191	124,097	44	300	4,400	29,200	69,500
	Placebo Survivors	34,942	123,514	43	300	4,300	29,400	69,200
	Real Deceased	37,449	126,254	44	300	4,600	29,900	99,500
	Placebo Deceased	37,691	125,537	43	300	4,500	30,000	98,900
Year of Co-inventor Death	Real Survivors	2006.629	3.42	2002	2004	2006	2009	2011
	Placebo Survivors	2006.723	3.44	2002	2004	2006	2009	2011
# Inventors	Real Deceased	4,714						
	Placebo Deceased	4,714						
	Real Survivors	14,150						
	Placebo Survivors	13,350						

Notes: This table presents summary statistics computed for the real and placebo deceased and survivor inventors. The statistics on number of co-inventors and firm size are computed in the year of death. The distribution of firm size is based on all inventors who receive a W2. For both real and placebo survivor inventors, about 10% of inventor-year observations are missing a W2, i.e. the inventors have no labor earnings (either because they are unemployed, self-employed or retired). Firm size is rounded to the nearest one hundred to preserve taxpayer confidentiality.

Table B.3.: Summary Statistics for Real and Placebo Coworkers and Second-Degree Connections

Variable	Sample	Mean	SD	10pc	25pc	50pc	75pc	90pc
	Real Second-degree Connections	175,247	358,347	46,000	81,000	116,000	170,000	267,00
Total Earnings	Placebo Second-degree Connections	174,900	350,102	45,000	82,000	115,000	173,000	266,000
	Real Coworkers	149,861	312,721	39,000	64,000	115,000	169,000	251,000
	Placebo Coworkers	154,627	316,266	40,000	65,000	118,000	174,000	254,000
	Real Second-degree Connections	144,449	291,697	39,000	70,000	108,000	156,00	239,000
Labor Earnings	Placebo Second-degree Connections	146,674	297,697	40,000	72,000	110,000	159,000	241,000
	Real Coworkers	114,559	257,233	22,000	56,000	91,000	142,000	200,000
	Placebo Coworkers	117,691	258,908	25,000	57,000	94,000	146,000	204,000
	Real Second-degree Connections	10.42	42.78	1	2	5	11	25
Cumulative Applications	Placebo Second-degree Connections	9.92	25.21	1	2	5	11	25
	Real Coworkers	2.31	2.51	0	1	1	3	7
	Placebo Coworkers	2.50	2.43	0	1	1	3	7
	Real Second-degree Connections	37.76	170.11	0.35	1.2	7	26.5	80.34
Cumulative Citations	Placebo Second-degree Connections	39.40	173.23	0.22	1.1	7.5	29.5	83
	Real Coworkers	6.64	12.2	0	0	1	6.58	23.5
	Placebo Coworkers	8.74	13.09	0	0	3	10	29.13
	Real Second-degree Connections	47.72	19.08	34	40	47	55	63
Age	Placebo Second-degree Connections	47.93	19.96	35	39	47	55	64
	Real Coworkers	44.28	12.94	30	36	44	52	59
	Placebo Coworkers	44.49	16.13	30	36	44	52	59
	Real Second-degree Connections	11,264						
# Inventors	Placebo Second-degree Connections	12,047						
	Real Coworkers	13,828						
	Placebo Coworkers	14,364						

Notes: This table reports summary statistics for the various groups of inventors defined in Section II.B, using data between 1999 and 2012 before the year of death. The table shows that the real and placebo second-degree connections and the real and placebo coworkers are very similar prior to co-inventor death, although our matching strategy did not use any information on these inventors. Note that the real and placebo second-degree connections are very similar to the survivor inventors, while the distribution of outcomes for real and placebo coworkers is very similar to that of the full sample. Dollar amounts are reported in 2012 dollars and are rounded to the nearest $1,000 to preserve taxpayer confidentiality. The balance between real and placebo coworkers and second-degree connections is qualitatively similar when considering the exact percentile values. For a detailed description of the data sources and sample construction, see Sections II.A and II.B.

Table B.4.: Balance for Number of Real and Placebo Survivor Coworkers per Deceased (Full Sample)

Variable	Sample	Mean	SD	10pc	25pc	50pc	75pc	90pc
Number of Inventor Coworkers	Real	52.38	100.61	1	4	19	63	143
In The Year of Death	Placebo	46.75	93.85	1	4	19	65	141
# Real Coworkers	143,646							
# Placebo Coworkers	173,128							

Notes: This table reports the number of real and placebo coworkers per real and placebo deceased inventor. There is good balance except in the tail, which creates an imbalance in the total number of real and placebo survivor coworkers.

B.2. Robustness Checks on the Causal Effect of Co-inventor's Premature Death

F-Test for Pretrending

We can formally test the hypotheses that the point estimates obtained by running specification (1) and shown in Figure 3 are all the same before and after co-inventor death, considering an equal number of periods before and after co-inventor death:

$H_0^{Before\ Death}: \ \beta_{-9}^{Real} = \beta_{-8}^{Real} = ... = \beta_{-2}^{Real}$

$H_0^{After\ Death}: \ \beta_{0}^{Real} = \beta_{2}^{Real} = ... = \beta_{7}^{Real}$

The results of the F-tests, shown in Table 2, confirm that there is no pretrending while there is an effect after death. We can reject at the 10% confidence level that all coefficients are similar after death for adjusted gross income and labor earnings, but we cannot do so for non-labor earnings and citations, which are more noisily estimated (although the point estimates reported in Figure 2 appear very stable). We can never reject that the point estimates are all similar before death. Appendix Table B3 tests for dynamic effects by pulling together several lags after death, which reduces noise.

Table B.5.: Testing For Dynamic Effects, P-Values of F-Tests

	Total Earnings	Labor Earnings	Non-Labor Earnings	Citation Count
For $H_0^{Before\ Death}$	0.671	0.875	0.690	0.764
For $H_0^{After\ Death}$	0.079	0.084	0.268	0.382

Notes: This panel reports the p-values of F-tests for equality of the β_k^{Real} coefficients from specification (1) before and after death, as specified by the hypotheses $H_0^{Before\ Death}$ and $H_0^{After\ Death}$ described in the text above the table. For more details on the outcome variables and the sample, see Table 2 and the main text. P-values are adjusted for the clustering of standard errors around the deceased inventors. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Balanced Panel

Table B.6.: Balanced Panel of Survivors Experiencing Co-inventor Death between 2003 and 2008

	Total Earnings	Labor Earnings	Labor Earnings >0	Patents Count	Citation Count
$AfterDeath^{Real}$	-2905.73**	-1907.36**	-0.0049*	-0.08090***	-0.0945***
s.e.	1345.88	806.25	0.00289	0.02957	0.0299
$AfterDeath^{All}$	199.025	-168.25	-0.00306**	-0.00622	-0.0293
s.e.	854.76	526.32	0.0021	0.02154	0.032
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	99,108	99,108	99,108	99,108	99,108
# Survivors	11,012	11,012	11,012	11,012	11,012
# Deceased	4,148	4,148	4,148	4,148	4,148
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2) on a balanced panel, keeping four years before and after death for each inventor in the sample. Specifically, we restrict the sample to survivor inventors whose associated deceased co-inventors passed away between 2003 and 2008 and we drop inventor-year observations when the lead or lag relative to co-inventor death is more than 4 years. Patent count is the number of patents the survivor inventor applied for in a given year, and citation count is the number of adjusted forward citations received on patents that the survivor applied for in a given year. Under the identification assumption described in Section III.B, β^{Real} gives the causal effect of co-inventor death on the various outcomes. The table shows that, for all outcome variables, we find a large and statistically significant effect. This indicates that the effect documented in Table 2 is not driven by the changing composition of the panel. The point estimates reported in this table are smaller than those reported in Table 2, because the balanced panel includes fewer inventor-year observations many years after death and Figure 3 shows that the negative effect on the survivors amplifies over time. Standard errors are clustered around the deceased inventors. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Dynamics

Table B.7.: Dynamic Causal Effect of Co-inventor Death, Full Sample

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-2,081**	-1,735**	-0.00658**	-0.0743***	-0.0939**
s.e.	(853)	(683)	(0.002712)	(0.0258)	(0.0375)
$AfterDeath^{Real} \cdot LongRun$	-2,949**	-1,990**	-0.00576**	-0.0504	-0.0507**
s.e.	(1,253)	(903)	(0.0026166)	0.0321	(0.0231)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

$$Y_{it=} \begin{array}{c} \beta^{Real} AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} + \widetilde{\beta^{Real}} AfterDeath_{it}^{Real} \cdot LongRun + \widetilde{\beta^{All}} AfterDeath_{it}^{All} \cdot LongRun \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where *LongRun* is an indicator equal to one for observations more than four years after death. The columns report the results for total earnings, labor earnings, employment, the count of patents and the count of citations. For all outcome variables, we find that the effect in the long run is significantly larger than in the short run following death events. For more details on the sample see Table 2 and the main text. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8.: Dynamic Causal Effect of Co-inventor Death, Sample Restricted to Deaths from 2003 to 2005

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-1,980**	-1,635**	-0.00558*	-0.0843***	-0.0839**
s.e.	(990)	(823)	(0.003112)	(0.0311)	(0.0412)
$AfterDeath^{Real} \cdot LongRun$	-2,743**	-2,001*	-0.00549**	-0.0404*	-0.0443*
s.e.	(1,365)	(1,103)	(0.002724)	(0.02421)	(0.02634)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	67,368	67,368	67,368	67,368	67,368
# Survivors	4,812	4,812	4,812	4,812	4,812
# Deceased	1,764	1,764	1,764	1,764	1,764
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

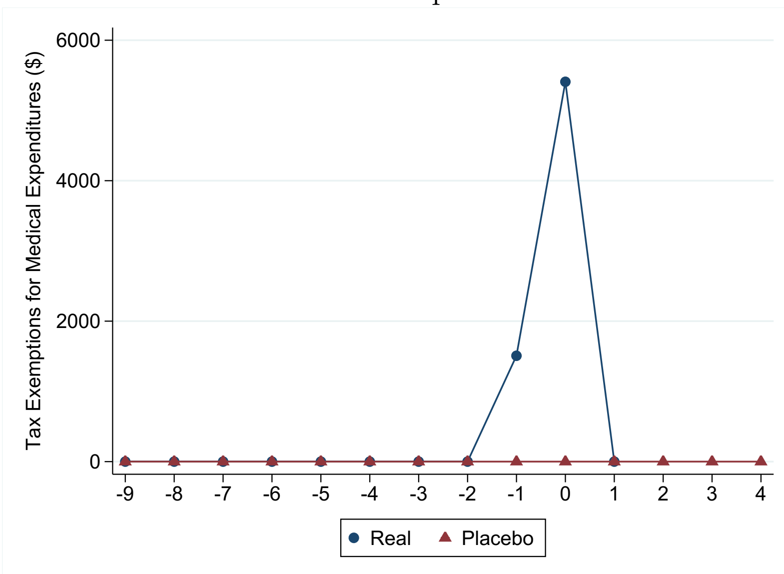
$$Y_{it=} \begin{array}{c} \beta^{Real} AfterDeath^{Real}_{it} + \beta^{All} AfterDeath^{All}_{it} + \widetilde{\beta^{Real}} AfterDeath^{Real}_{it} \cdot LongRun + \widetilde{\beta^{All}} AfterDeath^{All}_{it} \cdot LongRun \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where *LongRun* is an indicator equal to one for observations more than four years after death. The sample is restricted to the 4,812 co-inventors of the 1,764 real and placebo deceased with a year of death between 2003 and 2005. Inventor-year observations are dropped if the lag relative to co-inventor death is above seven years or if the lead relative to death is below four years. The various columns of the panel report the results for labor earnings, non-labor earnings, the count of patents and the count of citations. For all outcome variables, we find that the effect in the long run is significantly larger than in the short run following death events. The magnitude of the effects is similar to Figure 3 and Appendix Table B3, indicating that the dynamics of the effect are not driven by changes in the composition of the sample. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

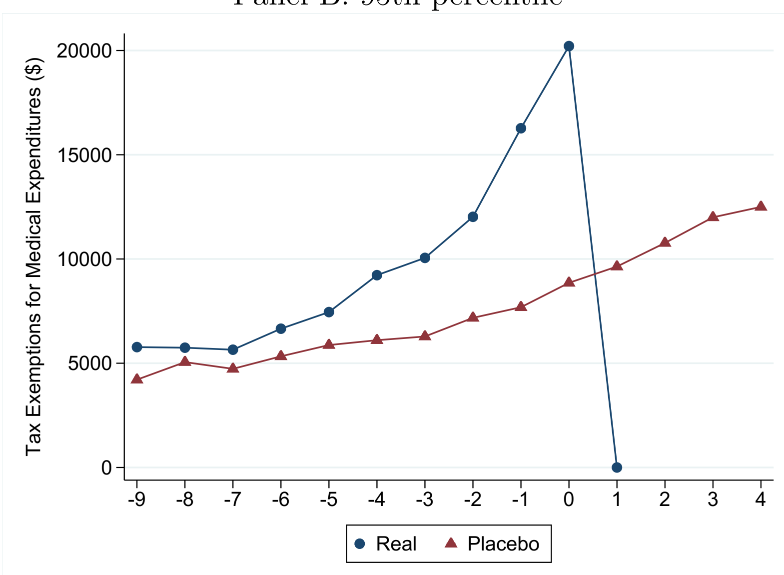
Anticipation

Figure B.2.: Tax Deductions for High Medical Expenditures Claimed by the Deceased

Panel A: 75th percentile



Panel B: 95th percentile



Notes: This figure shows the path of tax exemptions for medical expenditures claimed by the real and placebo deceased around the time of (real or placebo) death. For details on the sample, refer to Section II.B. Panel A shows that 75 percent of the real deceased inventors never claim any tax exemption for medical expenditures, except in the years just before death as well as during the year of death, suggesting that death is unanticipated for most survivors. Panel B shows that the 95th percentile of the distribution of tax deductions claimed for medical expenditures is very similar for real and placebo deceased until a few years before death, showing that some deaths result from lingering conditions and may therefore be anticipated.

Table B.9.: Results for Main Outcomes, Excluding Deceased who Claimed Any Tax Deduction for High Medical Expenditures

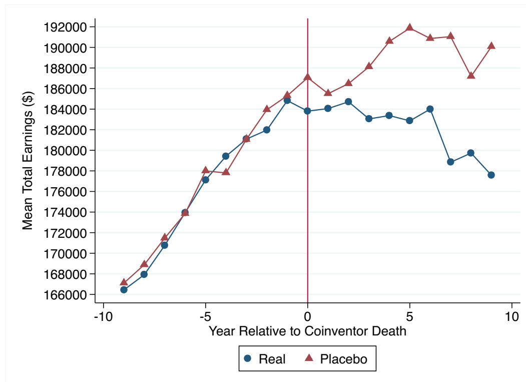
	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-4301.1562***	-3022.1***	-0.01047**	-0.1258***	-0.1017**
s.e.	1217.367	925.37	0.00417	0.0361	0.0442
$AfterDeath^{All}$	- 141.17	53.06	-0.00634**	-0.0020	0.0089
s.e.	576.10	595.30	0.0028	0.0231	0.00668
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	250,809	250,809	250,809	250,809	250,809
# Survivors	21,147	21,147	21,147	21,147	21,147
# Deceased	7,062	7,062	7,062	7,062	7,062
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2) in a sample that excludes all survivors whose associated deceased ever claimed tax deductions for medical expenditures. The table shows that the estimated causal effect of co-inventor death on the various outcomes is negative, statistically significant and large in magnitude. The point estimates are not very different but slightly larger than in Table 2. This result is not surprising, because our difference-in-differences estimator is biased downward if the causal effect of co-inventor impairement manifests itself before death. It bolsters the validity of the research design by showing that, if anything, we might be slightly underestimating the effect of co-inventor death due to lingering health conditions affecting some deceased inventors. Standard errors are clustered around the deceased inventors. $*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

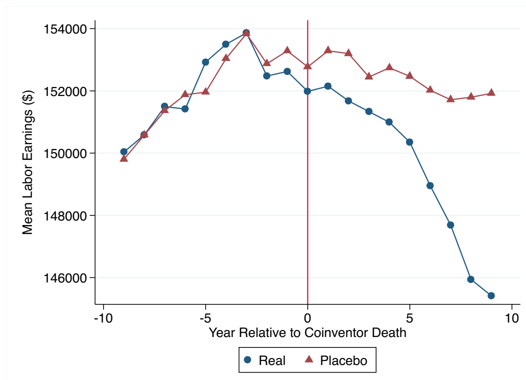
Alternative Matching Strategy

Figure B.3.: Path of Outcomes for Real and Placebo Survivor, Propensity Score Reweighting

Panel A: Survivor Inventors' Total Earnings



Panel B: Survivor Inventors' Labor Earnings



Panel C: Survivor Inventor's Adjusted Forward Citations Received for Patents Applied in Year

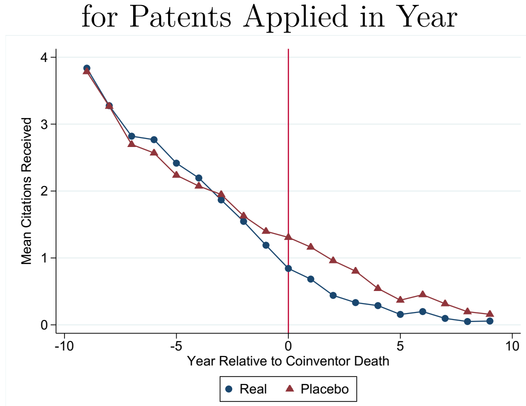


Table B.10.: The Causal Effect of Co-Inventor Death, Reweighting on the Propensity Score

	Total Earnings	Labor Earnings	Non-Labor Earnings	Patent Count	Citation Count
$AfterDeath^{Real}$	-3,624***	-2,621***	-1,032**	-0.0989***	-0.1103***
s.e.	(890)	(687)	(472)	(0.0236)	(0.0266)
$AfterDeath^{All}$	- 322	-51	552	-0.00081	0.07213
s.e.	(437)	(390)	(378)	(0.01452)	(0.12341)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	734,742	734,742	734,742	734,742	734,742
# Deceased	24,929	24,929	24,929	24,929	24,929
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and β^{All} from specification (2) in a sample of real and placebo survivors constructed following an alternative matching strategy, different from the one presented in the main text. Specifically, the matching strategy is as follows: (1) we identify all inventors who passed away before the age of 60 in our sample and we keep a random sample of 20,000 inventors who did not pass away during our sample ; (2) for each of the 20,000 inventors who did not pass away, we keep at random only one year of the sample, which will serve as our counterfactual year of death for these inventors in the following steps ; (3) we estimate the propensity score (which gives the probability of "treatment", i.e. the probability of passing away before the age of 60 between 1999 and 2012) by regressing an indicator for real deceased on age fixed effects, year of (real or placebo) death fixed effects, a fifth-order polynomial of wages in 1999, a fifth-order polynomial of total earnings in 1999, a fifth-order polynomial for cumulative patent applications at the time of death and a fifth-order polynomial for cumulative adjusted forward citations at the time of (real or placebo) death ; (4) we construct the co-inventor networks of all 24,929 real and placebo deceased in our sample for whom we have overlap in the propensity score ; (5) we run specification (2), which is described in the main text, in the sample of real and survivor inventor built in step (5) and using the propensity score estimated in step (2) as regression weight. The results reported in this table are very similar to the results reported in Table 2, showing that our results are robust to the choice of matching strategy. Note that the propensity-score reweighting strategy we employ here does not use any variable on the survivors, yet we find no pre-trending effects in Appendix Figure B2. Therefore, the details of the matching strategy do not matter for the substance of the results. It is important to use a matching strategy, however, because the real survivor inventors are in general older and of a higher level of achievement than the full sample of inventors, due to a selection effect (having a larger network of co-inventors increases the probability of experiencing the premature death of a co-inventor). For details about the outcome variables, refer to Table 2. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Citations

Table B.11.: Other Citation Metrics

	3-Year Citation Count Around Grant Year	5-Year Citation Count Around Grant Year	5-Year Examiner-Added Citation Count Around Grant Year
$AfterDeath^{Real}$	-0.095***	-0.1242***	-0.0943***
s.e.	(0.0245)	(0.0256)	(0.0342)
$AfterDeath^{All}$	0.135	-0.0739	0.086
s.e.	(0.1304)	(0.1345)	(0.1023)
Age and Year Fixed Effects	Yes	Yes	Yes
Individual Fixed Effects	No	No	No
# Observations	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428
Estimator	Poisson	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2), except that it does not include individual fixed effects because the Poisson estimator with individual fixed effects did not converge for several outcome variables. Appendix Table B8 shows that the results are similar with individual fixed effects, using a negative binomial estimator. The four outcome variables are as follows: (1) "3-year citation count around grant year" is the number of patents the survivor inventor applied for in a given year, weighted by the number of citations these patents received within three years of their respective year of grant; (2) "5-year citation count around grant year" is the number of patents the survivor inventor applied for in a given year, weighted by the number of citations these patents received within five years of their respective years of grant; (3) "5-year examiner-added citation count around grant year" is similar to the outcome variable in the second column, but taking into account only citations from patent examiners; (4) "5-year examiner-added citation count around grant year" is similar to the outcome variable in the second column, but taking into account only citations from applicants. For all outcome variables, we find a large and statistically significant effect. The magnitudes of these effects are similar to the effects reported in Table 2, Panel C, which shows the robustness of our result to the choice of the citation measure. For more details on the sample, see Table 2. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.12.: Citation Results with Negative Binomial Estimator and Individual Fixed Effects

	3-Year Citation Count Around Grant Year	5-Year Citation Count Around Grant Year	5-Year Examiner-Added Citation Count Around Grant Year	5-Year Applicant-Added Citation Count Around Grant Year
$AfterDeath^{Real}$	-0.09508***	-0.1291***	-0.1122***	-0.09636***
s.e.	0.0215	0.0312	0.03172	0.0297
$AfterDeath^{All}$	-0.1489***	-0.1691***	-0.161***	-0.1594***
s.e.	0.04621	0.04221	0.05231	0.04267
Age and Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes
# Observations	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428
Estimator	Negative Binomial	Negative Binomial	Negative Binomial	Negative Binomial

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2), using a negative binomial estimator. The five outcome variables are as follows: (1) "3-year citation count around grant year" is the number of patents the survivor inventor applied for in a given year, weighted by the number of citations these patents received within three years of their respective year of grant; (2) "5-year citation count around grant year" is the number of patents the survivor inventor applied for in a given year, weighted by the number of citations these patents received within five years of their respective years of grant; (3) "5-year examiner-added citation count around grant year" is similar to the outcome variable in the second column, but taking into account only citations added by patent examiners; (4) "5-year examiner-added citation count around grant year" is similar to the outcome variable in the second column, but taking into account only citations added by applicants; (5) citation count is the number of forward citations received on patents that the survivor applied for in a given year. For all outcome variables, we find a large and statistically significant effect. The magnitudes of these effects are similar to the effects reported in Table 2, Panel C, which shows the robustness of our results to the choice of estimator and the inclusion of individual fixed effects. For more details on the sample, see Table 2. Standard errors are clustered around the deceased inventors and computed by bootstrap with 100 draws. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Technology Classes

Table B.13.: Testing For Differences Across Technology Classes

	Total Earnings	Labor Earnings	Labor Earnings >0	Patents	Citations
$AfterDeath_{it}^{Real} \cdot Tech1$	-3,883*	-2,200*	-0.0075*	-0.0701**	-0.1065**
s.e.	(2,273)	(1,135)	(0.0044)	(0.0305)	(0.04875)
$AfterDeath_{it}^{Real} \cdot Tech2$	-4,208**	-2,710**	-0.0096*	-0.1406 ***	-0.1234***
s.e.	(2,054)	(1,319)	(0.0049)	(0.0440)	(0.0395)
$AfterDeath_{it}^{Real} \cdot Tech3$	-4,505*	-3,462***	-0.0063*	-0.092***	-0.1180***
s.e.	(2,364)	(1,333)	(0.0038)	(0.0341)	(0.0413)
$AfterDeath_{it}^{Real} \cdot Tech4$	-3,498**	-2,507*	-0.0117**	-0.1021*	-0.0954*
s.e.	(1,613)	(1,331)	(0.00518)	(0.0556)	(0.05096)
$AfterDeath_{it}^{Real} \cdot Tech5$	-3,080*	-2,075*	-0.0086*	-0.0692**	-0.0743*
s.e.	(1,740)	(1,102)	(0.0047)	(0.0343)	(0.0389)
$AfterDeath_{it}^{Real} \cdot Tech6$	-4,402*	-3,233**	-0.0048*	-0.064**	-0.072**
s.e.	(2,476)	(1,314)	(0.0028)	(0.0292)	(0.0312)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
F-Test on Equality of All β_{TechT}^{Real}	0.62	0.45	0.42	0.38	0.51
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients $\beta_{Tech\ T}^{Real}$ from the following specification:

$$Y_{it} = \begin{array}{c} \beta^{Real} AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} \\ + \sum_{T=1}^{6} \widetilde{\beta_{TechT}^{Real}} AfterDeath_{it}^{Real} \cdot TechT + \sum_{T=1}^{6} \widetilde{\beta_{TechT}^{All}} AfterDeath_{it}^{All} \cdot TechT \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where $TechT$ is an indicator equal to one when a survivor inventor has invented most of his patent prior to the year of co-inventor death in technology class T (we aggregate USPC classes into six main technology classes, as in Hall *et al.*, 2001). The distribution of real and placebo survivor inventors across the six main technology classes we consider is presented in Appendix Table A1. Technology class #1 is Chemical, #2 is Computers and Communications, #3 is Drugs and Medical, #4 is Electrical & Electronic, #5 is Mechanical and #6 is Others. The point estimates show significant effects for all outcomes in all technology classes, indicating that our results are not driven by a particular technology class. Formally, for each outcome we report the p-value of a F-test for the hypothesis:

$$H_0 : \beta_{Tech1}^{Real} = \beta_{Tech2}^{Real} = ... = \beta_{Tech6}^{Real}$$

We fail to reject that the effect is the same across all technology classes. We have investigated the robustness of these results by running regressions in subsamples, considering in turns populations of survivor inventors specializing in each of the six technology classes before the year of co-inventor death. The results are qualitatively similar. For details on the sample, see Table 2. Standard errors are clustered around the deceased inventors and the p-values of F tests are adjusted accordingly. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Inference Accounting for the Matching Step

Table B.14.: Inference on The Causal Effect of Co-Inventor Death Accounting For the Matching Step

	Total Earnings	Labor Earnings	Labor Earnings >0	Non-Labor Earnings	Patents	Citations
$AfterDeath^{Real}$	-3,875***	-2,720***	-0.00914***	-1,199**	-0.0916***	-0.092***
s.e.	(839)	(659)	(0.00288)	(473)	(0.0178)	(0.0214)
$AfterDeath^{All}$	-215	-38	-0.0049**	652*	0.0006	0.0508
s.e.	(529)	(451)	(0.0021)	(357)	0.0182	0.1161
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	No
# Observations	325,726	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500	27,500
# Matched Pairs	4,714	4,714	4,714	4,714	4,714	4,714
Estimator	OLS	OLS	OLS	OLS	Poisson	Poisson

Notes: This table reports the estimated coefficients β^{Real} and β^{All} from specification (2). For details about the outcome variables and the sample, refer to Table 2. The difference between this table and Table 2 is that, here, standard errors are computed using the "coupled bootstrap" procedure presented in Abadie and Spiess (2015). We use one hundred bootstrap replications for each of the six outcome variables and we have checked that the results are similar when bootstrapping one thousand times for total earnings. The coupled bootstrap method applied to our setting works as follows: one redraws with replacement *pairs* of matched real-placebo deceased and all of their associated survivors (i.e. the full panel of observations for all of these survivors). The coupled bootstrap is effectively just a block bootstrap, but we re-sample together treated and matched control units, which reflects the dependency between treated and matched control units through the matched covariates (in our setting, the treated and matched control units are the real and placebo deceased). In contrast, in the standard bootstrap, treated and control units are treated as independent and are not resampled togethed. Note that the validity of the coupled bootstrap follows from a general result that applies to smooth functionals of the marginal outcome distributions, therefore it should be valid for inference on the difference-in-differences specification we run in our sample of real and placebo survivor inventors. The standard errors we obtain through this procedure are slightly smaller than the clustered standard errors reported in Table 2, which shows the robustness of our results. These smaller standard errors may result from a high positive correlation between the potential outcomes conditional on covariates, which is reasonable in our setting. Refer to Abadie and Spiess (2015) for more details.

B.3. Additional Results on Mechanism

What Does the Reduced-form Effect of an Inventor's Death Imply about Complementarity and Substitutability Patterns between Inventors?

Consider a survivor inventor and a prematurely deceased inventor who used to be co-inventors, coworkers, or part of an extended co-inventor network. As mentioned in the main text of the paper, our quasi-experiment does not deliver insights about general substitution and complementarity patterns between these inventors. The reduced-form effects we identify correspond to the idiosyncratic effect of an inventor on their co-inventors, coworkers and second-degree connections. Formally, the sign of our reduced-form coefficients identifies substitutability and complementarity patterns between two inventors conditional on irreplaceability. A non-zero point estimate rejects the null that all of the tasks performed by the prematurely deceased inventor were perfectly replaceable (i.e. it is not possible for the surviving inventor to find another inventor playing the exact same role as the deceased inventor). However, we cannot reject that at least some of the tasks performed by the deceased were replaceable. The sign of the point estimate for the effect of inventor death on the various outcomes of interest reflects complementarity and substitutability patterns for the tasks performed by the prematurely deceased inventors that were not replaceable, and only for those tasks. Specifically, a positive (negative) point estimate tells use that those tasks were on average substitutable for (complementary with) the tasks performed by the survivor inventor. In contrast, we do not learn about complementarity and substitutability patterns for the tasks performed by the deceased that were replaceable.

The Nature of Team-Specific Capital: Match vs. Experience

Team-specific capital can result from a "match" component which is constant over time or from an "experience" component which increases the value of the collaboration over time. The idiosyncratic value of a collaborative relationship may also vary over the lifecycle of an inventor, e.g. if it is more difficult to substitute for co-inventors later in life. From the point of view of inventor i, one could thus conceptualize the idiosyncratic value of a collaborative relationship with inventor j at time t (denoted V_{ijt}) as resulting from a match component θ_0^{ij}, an experience component θ_1 and lifecycle covariates X_{it}:

$$V_{ijt} = \theta_0^{ij} + \theta_1(t - T_{ij}) + \gamma X_{it} + \epsilon_{ijt}$$

where T_{ij} denotes the time of the first collaboration between i and j.

To separately identify θ_0^{ij} and θ_1 in this simple empirical model, the ideal experiment would follow three steps: randomly assign inventors to work in teams; separate the teams after t years of collaboration, where t varies randomly across teams; test whether the loss in output if larger for teams that were separated later, controlling for inventor age at separation. This ideal empirical design can be approximated in our setting, using the difference between the year of co-inventor death and the year of first collaboration as a measure of "potential length of collaboration", which could serve as an instrument for the actual length of the collaboration between the two inventors. Note that

$$PotentialCollaborationLength_{ij} \equiv YearCoinventorDeath_{ij} - YearFirstCollaboration_{ij} = AgeAtCoinventorDeath_i - AgeAtFirstCollaboration_i$$

In our non-experimental setting, the formation of teams is endogenous and, therefore, the age at first collaboration could be correlated with match quality θ_0^{ij} (e.g. if inventors who think alike and were trained in the same schools are more likely to meet earlier in life). Because of the collinearity between potential collaboration length and age effects shown in the equation above, we cannot control for both age at first collaboration and age at co-inventor death.

Given this limitation, we leave an in-depth study of the match and experience components of team-specific capital to future research.

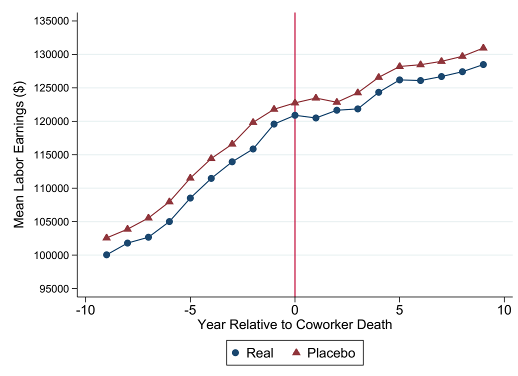
However, our setting can be used to provide suggestive evidence on the relative importance of the match and experience components of team-specific capital. First, we document heterogeneity in the treatment effect depending on the potential length of collaboration between the two inventors while controlling for the survivor's age at death. We find that the potential length of collaboration is positively correlated with the magnitude of the treatment effect, which provides suggestive evidence for the importance of the experience component.[1] Second, we show in Table C6 that firm size is not an important predictor of treatment effect heterogeneity. Since it is likely that it is easier for an inventor to find new collaborators in a larger firm, this result provides suggestive evidence that the match component may not be the primary determinant of the idiosyncratic value of a collaboration.

Our results have established that team-specific capital has a first-order impact in an inventor's career, therefore it would be worthwhile to distinguish between the match and experience components in future research. A promising direction would be to develop a suitable experimental design. Another promising direction would be to estimate a structural model using observational data on repeated collaborations between patent inventors and test the key prediction of the "experience" view, namely that the quality of a collaboration should increase over time (conditional on inventor fixed effects, lifecycle controls and other covariates).

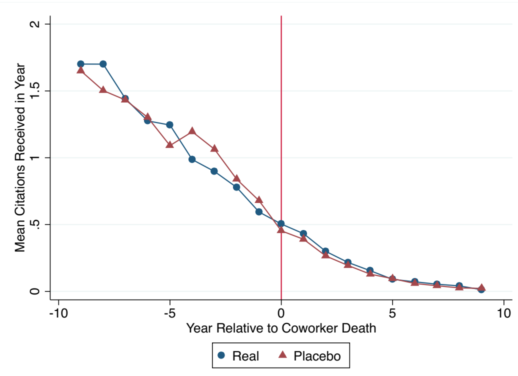
[1] Of course, heterogeneity in the treatment effect by actual length of collaboration, as reported in Table 7, doesn't allow us to distinguish between the match and experience components because teams with a high match quality will endogenously collaborate for a longer period of time. The results for heterogeneity by potential length of collaboration are available from the authors upon request.

Figure B.4.: Path of Outcomes for Coworkers and Second-Degree Connections around Year of Death

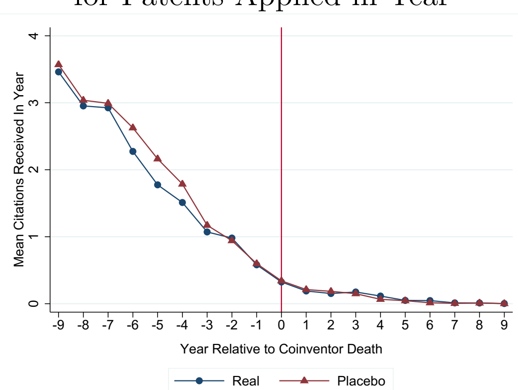
Panel A: Coworkers' Labor Earnings



Panel B: Coworker's Adjusted Forward Citations Received for Patents Applied in Year



Panel C: Second-degree Connections' Adjusted Forward Citations Received for Patents Applied in Year



Causal Effect of Coworker Death in the Full Sample

Table B.15.: Causal Effect of Coworker Death, Including Coworkers in Firms of Any Size

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
β^{Real}	105.21	336.05	0.0034	0.0149	0.0048
s.e.	(461.22)	(312.59)	(0.0048)	(0.0110)	(0.0041)
β^{All}	-521	-702.5	-0.004357*	-0.0366**	-0.00623*
s.e.	(518)	(653)	(0.00241)	(0.01462)	(0.00355)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	3,642,901	3,642,901	3,642,901	3,642,901	3,642,901
# Coworkers	316,774	316,774	316,774	316,774	316,774
# Deceased	6,289	6,289	6,289	6,289	6,289
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and β^{All} from specification (2) for the sample of coworkers, considering deceased inventors in firms of any size. The five outcome variables are as follows: (1) total earnings; (2) labor earnings; (3) an indicator equal to one when the inventor receives a W-2, i.e. is employed; (4) the number of patents the coworker applied for in a given year; (5) the number of forward citations received on patents that the coworker applied for in a given year (therefore, this variable reflects the timing and quality of patent applications by the survivor, not the timing of citations). Under the identification assumption described in Section III.B, β^{Real} gives the causal effect of coworker death on these various outcomes. We do not find any significant effect for any of the outcomes, and the point estimates are positive. These results are qualitatively similar to those presented in Table 3: the absence of a negative effect on coworkers rules out the theory that the large effects documented in Section III are driven by the disruption of the firm. In contrast with Table 3, we no longer find positive and significant effects on the extensive margin of labor earnings, patents and citations, which could be because the firms we consider here are too large for any substitutability pattern to operate between inventor coworkers on average. Inventor-year observations are dropped when the lead or lag relative to coworker death is above 9 years. The unbalanced nature of this panel is the same for real and placebo coworkers. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. $*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Sample Sizes for Results by Relative Ability Levels

Table B.16.: Sample Sizes for Analysis by Relative Ability Levels

Deceased / Survivor Earnings Quartile	1	2	3	4
1	42,431 / 4,040 / 2,706	22,300 / 1,884 / 1,132	1,9619 / 1,706 / 1,062	17,251 / 1,456 / 887
2	20,968 / 1,747 / 1,150	37,390 / 3,382 / 1,625	28,158 / 2,485 / 1,349	17,476 / 1,506 / 975
3	20,085 / 1,685 / 989	15,899 / 1,366 / 617	20,465 / 1,686 / 711	11,696 / 1,071 / 549
4	9,132 / 825 / 354	11,090 / 981 / 379	11,540 / 1053 / 477	14,354 / 1,313 / 535

Notes: This panel reports the sample sizes for each of the sixteen subsamples studied in the various panels of Table 5. Each of these subsamples corresponds to a different combination for the total earnings quartiles of the survivor and the deceased. The earnings quartiles are computed three years before death. Within each cell, the sample sizes are reported according to the following format: Number of observations / Number of survivors / Number of deceased. For instance, in the subsample of survivor inventors who were in the lowest earnings quartile three years before death and whose associated deceased was also in the lowest earnings quartile at that time, we have 2,706 real and placebo deceased, 4,040 real and placebo survivors, and 42,432 inventor-year observations.

Probability of Changing Firms

Table B.17.: Causal Effect of Co-Inventor Death on the Probability of Changing Firm

	Changing Firm
$AfterDeath_{it}^{Real}$	-0.00124
s.e.	(0.00192)
$AfterDeath_{it}^{All} \cdot SmallFirm$	0.00798**
s.e.	(0.004016)
Age and Year Fixed Effects	Yes
Individual Fixed Effects	Yes
# Observations	266,087
# Survivors	22,740
# Deceased	8,382
Estimator	OLS

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

$$ChangingFirm_{it} = \begin{array}{l} \beta^{Real} AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} \\ +\widetilde{\beta^{Real}} AfterDeath_{it}^{Real} \cdot SmallFirm + \widetilde{\beta^{All}} AfterDeath_{it}^{All} \cdot SmallFirm \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

where (1) $ChangingFirm_{it}$ is an indicator variable equal to 1 if the deceased is employed in a different firm in year t compared with the year prior to co-inventor death; (2) $SmallFirm$ is an indicator equal to one if the survivor was in a firm with less than one hundred employee in the year prior to coinventor death; (3) the rest of the specification is similar to specification (2) in the main text. The table shows that in general co-inventor death does not have a statistically significant impact on an inventor's probability of changing firms. However, survivor inventors who are in a small firm are more likely to change firms after co-inventor death. This finding is consistent with the view that the survivor inventor may be looking for new co-inventors and may change firms to do so. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Probability of Getting a New Co-inventor

Table B.18.: Causal Effect of Co-Inventor Death on the Probability of Getting a New Co-inventor

	New Co-Inventor In Year
β^{Real}	0.05899
s.e.	(0.067409)
β^{All}	-0.107534*
s.e.	(0.060466)
Age and Year Fixed Effects	Yes
Individual Fixed Effects	Yes
# Observations	325,726
# Survivors	27,500
# Deceased	9,428
Estimator	OLS

Notes: This panel reports the estimated coefficients β^{Real} and β^{All} for specification (2), using as an outcome variable the number of new coinventors of the survivor in a given year. This variable is built using data on patent applications and counts the number of new co-inventors of the survivor in a given year, i.e. the number of inventors who apply for a patent with the survivor in this year and who had never applied for a patent with the survivor in any of the previous years. We find no statistically significant effect, and the point estimate is small in magnitude. This suggests that the survivor inventor is not able to find substitutes for the deceased co-inventor, which may explain the strength of the effect on the survivor's earnings and patents documented in Table 2. Note that the outcome variable in this table is not a perfect measure of changes in collaboration patterns, since it is based on patent applications, i.e. we can observe the new co-inventor only when a patent application is filed. This creates a censoring problem, which however is similar for treated and control inventors. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity by Survivor's Age

Table B.19.: Heterogeneity in Causal Effect of Co-Inventor Death by Age Quartile

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-3,484***	-2,526***	- 0.00476	-0.09781***	-0.10962***
s.e.	(1,102)	(724)	(0.00312)	(0.02915)	(0.03451)
$AfterDeath^{Real} \cdot AgeQ2$	33	-218	0.00014	-0.00385	0.02808
s.e.	(549)	(412)	(0.00088)	(0.0046)	(0.03602)
$AfterDeath^{Real} \cdot AgeQ3$	-990	-149	-0.00451**	0.001311	-0.00129
	(950)	(567)	(0.00208)	(0.04823)	(0.00314)
$AfterDeath^{Real} \cdot AgeQ4$	-1,533	-1,011	-0.00964***	-0.0498*	-0.00535
	(1,288)	(738)	(0.00352)	(0.02959)	(0.00371)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}_{Qk}}$ from the following specification:

$$Y_{it} = \beta^{Real} AfterDeath^{Real}_{it} + \beta^{All} AfterDeath^{All}_{it} + \sum_{k=2}^{4} \widetilde{\beta^{Real}_{Qk}} AfterDeath^{Real}_{it} \cdot AgeQk + \sum_{k=2}^{4} \widetilde{\beta^{All}} AfterDeath^{All}_{it} \cdot AgeQk + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it}$$

using similar notation to Section III.B and where $AgeQk$ is an indicator equal to one when the survivor is in the k-th quartile of age at co-inventor death. The specification with the Poisson estimator for columns 4 and 5 of the table is similar. The table shows that there is no significant heterogeneity in the causal effect of co-inventor death on the various outcomes by age quartile, except on the extensive margin of labor earnings, where the effect is driven by survivors who were older at the time of co-inventor death. For younger survivor inventors, the point estimate for the effect on the extensive margin of labor earnings is an imprecisely estimated zero. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity by Firm Size

Table B.20.: Heterogeneity in Causal Effect of Co-Inventor Death by Firm Size Quartile

	Total Earnings	Labor Earnings	Labor Earnings >0	Patent Count	Citation Count
$AfterDeath^{Real}$	-3,506***	-2,537***	-0.0094**	-0.0989***	-0.1020 ***
s.e.	(878)	(690)	(0.0041)	(0.0245)	(0.0234)
$AfterDeath^{Real} \cdot FirmQ2$	-422	169	0.0008	0.0012	0.0023
s.e.	(633)	(587)	(0.0013)	(0.0093)	(0.0036)
$AfterDeath^{Real} \cdot FirmQ3$	-395	-365	-0.0003	-0.0123	0.0032
	(533)	(453)	(0.0021)	(0.0187)	(0.0092)
$AfterDeath^{Real} \cdot FirmQ4$	198	-204	-0.0023	0.0021	0.0182
	(643)	(346)	(0.0017)	(0.0163)	(0.015)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	284,707	284,707	284,707	284,707	284,707
# Survivors	23,925	23,925	23,925	23,925	23,925
# Deceased	8,768	8,768	8,768	8,768	8,768
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}_{Qk}}$ from the following specification:

$$Y_{it} = \beta^{Real} AfterDeath^{Real}_{it} + \beta^{All} AfterDeath^{All}_{it} + \sum_{k=2}^{4} \widetilde{\beta^{Real}_{Qk}} AfterDeath^{Real}_{it} \cdot FirmQk + \sum_{k=2}^{4} \widetilde{\beta^{All}} AfterDeath^{All}_{it} \cdot FirmQk + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it}$$

using similar notation to Section III.B and where $FirmQk$ is an indicator equal to one when the survivor is in the k-th quartile of firm size in the year of co-inventor death. The specification with the Poisson estimator for columns 4 and 5 of the table is similar. The table shows that there is no significant heterogeneity in the causal effect of co-inventor death on the various outcomes by firm quartile. The sample includes all real and placebo survivor inventors who received a W2 at the time of co-inventor death. Inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity by Citizenship Status

Table B.21.: Heterogeneity in Causal Effect of Co-Inventor Death by Survivor's Citizenship Status

	Total Earnings	Labor Earnings	Labor Earnings>0	Patent Count	Citation Count
$AfterDeath^{Real}$	-3,675***	-2,604***	-0.0982***	-0.079***	-0.1056***
s.e.	(918)	(683)	(0.0328)	(0.0243)	(0.0271)
$AfterDeath^{Real} \cdot Foreigner$	-727	-506	0.0083	-0.0463 **	0.0263
s.e.	(663)	(421)	(0.0098)	(0.0214)	(0.0209)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No
# Observations	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	Poisson	Poisson

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

$$Y_{it} = \begin{array}{c} \beta^{Real} AfterDeath^{Real}_{it} + \beta^{All} AfterDeath^{All}_{it} + \widetilde{\beta^{Real}} AfterDeath^{Real}_{it} \cdot Foreigner + \widetilde{\beta^{All}} AfterDeath^{All}_{it} \cdot Foreigner \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where *Foreigner* is an indicator turning to one when the survivor inventor is not a US citizen. The table shows that there is no significant heterogeneity in the causal effect of co-inventor death by citizenship status, except for patent count. This result is consistent with the notion that it may be more difficult for foreign inventors to find new co-inventors, hence a stronger decline in citations, but at the same time they may not be rewarded for performance on the same basis as US inventors, explaining the absence of differential effect on earnings. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Heterogeneity by Network Size

Table B.22.: Heterogeneity in Causal Effect of Co-Inventor Death by Survivor's Network Size

	Total Earnings	Labor Earnings	Labor Earnings> >0)	Patents	Citations	New Co-inventor
β^{Real}	-3,573***	-2,615***	-0.0095***	-0.0891***	-0.0952***	0.0239
s.e.	(857)	(706)	(0.0034)	(0.0237)	(0.0232)	(0.0632)
$\beta^{Real} \times Small\ Network$	-534	-283	0.0012	-0.0057	0.0067	0.0884
s.e.	(614)	(450)	(0.0023)	0.0102	(0.0192)	(0.059)
Age and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	No	No	Yes
# Observations	325,726	325,726	325,726	325,726	325,726	325,726
# Survivors	27,500	27,500	27,500	27,500	27,500	27,500
# Deceased	9,428	9,428	9,428	9,428	9,428	9,428
Estimator	OLS	OLS	OLS	OLS	Poisson	OLS

Notes: This panel reports the estimated coefficients β^{Real} and $\widetilde{\beta^{Real}}$ from the following specification:

$$Y_{it} = \begin{array}{c} \beta^{Real} AfterDeath_{it}^{Real} + \beta^{All} AfterDeath_{it}^{All} \\ +\widetilde{\beta^{Real}} AfterDeath_{it}^{Real} \cdot SmallNetwork + \widetilde{\beta^{All}} AfterDeath_{it}^{All} \cdot SmallNetwork \\ + \sum_{j=25}^{70} \lambda_j 1_{\{age_{it}=j\}} + \sum_{m=1999}^{2012} \gamma_m 1_{\{t=m\}} + \alpha_i + \epsilon_{it} \end{array}$$

using similar notation to Section III.B and where $SmallNetwork$ is an indicator turning to one when the size of the co-inventor network of the survivor inventor is below median at the time of death. The table shows that there is no significant heterogeneity in the causal effect of co-inventor death by network size. This result is qualitatively similar when considering other interaction terms (linear, quartile) based on survivor's network size at the time of death. An explanation for this finding is that the observed network of co-inventors at the time of death may be a noisy proxy for the survivor's actual network, given that collaborations are ongoing before patent applications are filed. Overall, the network size variable appears to be a less reliable indicator of the difficulty for the survivor to recover from the death of his co-inventor than the measures of collaboration intensity presented in Table 6. The sample includes all real and placebo survivor inventors in a 9-year window around the year of co-inventor death, i.e. inventor-year observations are dropped when the lead or lag relative to co-inventor death is above 9 years. The unbalanced nature of this panel is the same for real and placebo inventors. Dollar amounts are reported in 2012 dollars. Standard errors are clustered around the deceased inventors. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.4. Econometric Considerations

What is Identified In Specification (1)?

This appendix considers specification (1) introduced in Section III and asks what is identified about the coefficients $\{\beta^{Real}(k)\}$ and $\{\beta^{All}(k)\}$. k denotes the year relative to co-inventor death, which can be expressed as the difference between the time of co-inventor death (CDT_i) and time τ (so $k = \tau - CDT_i$). We delay imposing any "normalization" on the model and we note that $\forall \mu \in R$:

$$\begin{aligned} \beta^{All}(\tau - CDT_i) + \gamma(\tau) + \alpha(i) &= [\beta^{All}(\tau - CDT_i) - \mu(\tau - CDT_i)] + [\gamma(\tau) + \mu \cdot \tau] + [\alpha(i) - \mu \cdot CDT_i] \\ &= \widetilde{\beta^{All}}(\tau - CDT_i) + \widetilde{\gamma}(\tau) + \widetilde{\alpha}(i) \end{aligned}$$

Therefore, any function of the full vector coefficients, $G(\beta^{All}(.))$, is not identified unless $G(\beta^{All}(.) + h(.)) = G(\beta^{All}(.))$ for any linear function $h(k) = \alpha_1 + \alpha_2 k$. This observation helps understand which predictive effects are identified.[2] If $G(\beta^{All}, \gamma, \alpha)$ is identified, then we can evaluate it and we will get a well-defined predicted value. In specification (1), any solution to the least-squares fit gives the same value for $G(\beta^{All}, \gamma, \alpha)$. Although the solution of the least-square fit in specification (1) is not unique because the regressor matrix does not have full column rank, there is a unique predicted value.

The intuition for this result is that the set of leads and lags associated with $\beta^{All}(k)$ applies to all individuals in the sample. As a result, when we first-difference the data to eliminate the individual fixed effects, we lose information about a linear trend that could affect all individuals either through the $\beta^{All}(k)$ coefficients or through the year or age fixed effects. So $\beta^{All}(k)$, the age fixed effects and the year fixed effects are identified only up to a linear time trend. In practice, when estimating specification (1), we can drop any two dummies within the set of age or year with fixed effects or within the set of leads and lags $\beta^{All}(k)$. This will serve as our "normalization" for the linear trend.

In contrast, $\beta^{Real}(k)$ is associated with a set of leads and lags that can turn to one only for the real survivors. As a result, $\beta^{Real}(k)$ is identified up to a level shift affecting all coefficients. Due to the individual fixed effects, one of the $\beta^{Real}(k)$ must be normalized to zero, as is usually the case in estimators with a full set of leads and lags around an event.

[2]The point of a "normalization" is that imposing it will not affect the value of a predictive effect that is identified: to be identified means identified without any normalization.

Empirical Relevance

Our specifications (1) and (2) are an application of the standard difference-in-differences estimator to our setting. The current practice in the literature with a setting similar to ours, for instance Azoulay *et al.* (2010) and Oettl (2012), is to use specifications including age, year and individual fixed effects only, without including L_{it}^{All} (as in specification (1)) or $AfterDeath_{it}^{All}$ (as in specification (2)). Becker and Hvide (2013) present a specification similar to our specification (2), but appropriately testing for pre-trending requires using specification (1), as we do.

The point that age, year and individual fixed effects may not fully account for trends in life-time earnings and patents around co-inventor death is a simple but crucial one. Had we not included $AfterDeath_{it}^{All}$ in specification (2), we would have over-estimated the effect of co-inventor death on the probability of being employed by 50% (Table 2, Panel B), we would have spuriously concluded that an inventor death causes a decline in the patents and in the probability of being employed of this inventor's coworkers and second-degree connections (Table 3, Panels A and B), and we would have mistaken mean-reversion patterns for heterogeneity in the causal effect of co-inventor death by relative ability level of the survivor and the deceased (Table 5, Panels B and C).

B.5. Data Appendix

This section documents the most important steps for the construction of the matched inventor-taxpayer database from Bell et al. (2015), provides a comparison of the distribution of Census firm size and EIN size, and gives summary statistics on the composition of patent inventor teams.

A. Data Construction

A.1 Data Preparation

- **Suffix Standardization**. Suffixes may appear at the end of taxpayers' first, middle, or last name fields. Any time any of these fields ends with a space followed by "JR", "SR", or a numeral I-IV, the suffix is stripped out and stored separately from the name[3].

- **First name to imputed first/middle name**. The USPTO separates inventor names into "first" and "last," but the Treasury administrative tax files often separate names into first, middle, and last. In practice, many inventors do include a middle initial or name in the first name field. Whenever there is a single space in the inventor's first name field, for the purposes of matching, we allow the first string to be an imputed first name, and the second string to be an imputed middle name or initial. The use of these imputed names is outlined below.

A.2 Pseudo code for Match on Name and Location

The exact matching stages are as follows. We conduct seven progressive rounds of matching. Inventors enter a match round only if they have not already been matched to a taxpayer in an earlier round. Each round consists of a name criterion and a location criterion. The share of data matched in each round is noted, with an impressive 49% being exact matches on the first stage.

- The matching algorithm takes as input a relation of inventor data and five relations of Treasury administrative tax files:
  - Input relations:
    - Inventors(inv_id, first, last, imputed_first, imputed_middle, suffix) - directly from USPTO

[3] Numerals I and V are only permissive suffixes at the end of a last name field, as these may be middle initials in a middle name field.

    - NamesW2(irs_id, first, middle, last, suffix) - all names used by individual on W2 information returns; name field is recorded as first, middle, and last
    - Names1040(irs_id, first, middle, last) - all self-reported names from 1040 forms[4]
    - Nameln1W2(irs_id, fullname) - all names from W2, but a separate variable not recorded as first, middle, last that was more frequently present
    - CitiesW2(irs_id, city, state) - all cities reported on W2
    - Zips1040(irs_id, name) - all zip codes reported on 1040
  - Output relation:
    - Unique-Matches (inv_id, irs_id)

- **Stage 1**: Exact match on name and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. Either the inventor's first name field exactly matches the concatenation of the Treasury administrative tax files first and middle name fields or the Treasury administrative tax files middle name field is missing, but the first name fields match. If an imputed middle name is available for the inventor, candidate matches are removed if they have ever appeared in Treasury administrative tax files with a middle name or initial that conflicts with the inventor's.
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 49% of patents are uniquely matched in this stage.
- **Stage 2**: Exact match on imputed name data and location.

[4] We only take names off of 1040s for those who file singly because it proved difficult to parse names of those list them jointly

- Name match: The inventor's last name exactly matches the taxpayer's last name and the taxpayer's last name is the same as the inventor's imputed first name. Either the inventor's imputed middle name/initial matches one of the taxpayer's middle/initial name fields, or one of the two is missing. For inventors with non-missing imputed middle names, priority is given to matches to correct taxpayer middle names rather than to taxpayers with missing middle names. As above, candidate matches are removed if they have ever appeared in Treasury administrative tax files with a conflicting middle name or initial.
- Location match: As above, the inventor's city and state must match some city and state reported by that taxpayer exactly.
- 12% of patents are uniquely matched in this stage.

• **Stage 3**: Exact match on actual or imputed name data and 1040 zip cross-walked.

- Name match: The inventor's last name exactly matches the taxpayer's last name. The inventor's first name matches the taxpayer's first name in one of the following situations, in order of priority:

1. Inventor's firstname is the same as the taxpayer's combined first and middle name.
2. Inventor's imputed firstname matches taxpayer's and middle names match on initials.
3. The inventor has no middlename data, but inventor's firstname is the same as the taxpayer's middle name.

- As always, taxpayers are removed if they are ever observed filing with middle names in conflict with the inventor's.
- Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.

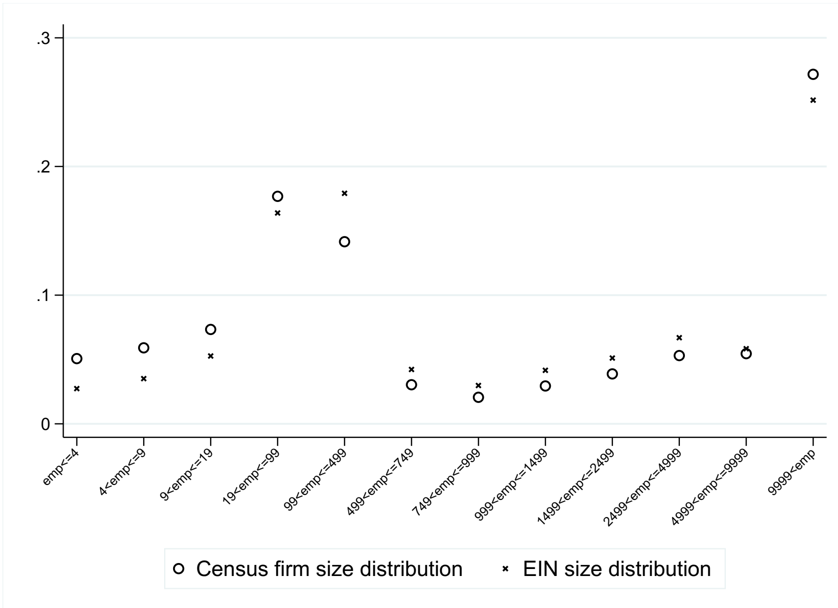
- 3% of patents are uniquely matched in this stage.

- **Stage 4**: Same as previous stage, but using 1040 names instead of names from W2's.
  - Name match: The inventor's name matches the name of a 1040 (or matches without inventor's middle initial/name, and no taxpayer middle initials/names conflict with inventor's).
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 6% of patents are uniquely matched in this stage.
- **Stage 5**: Match using W2 full name field.
  - Name match: The inventor's FULL name exactly matches the FULL name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 8% of patents are uniquely matched in this stage.
- **Stage 6**: Relaxed match using W2 full name field.
  - Name match: The inventor's full name (minus the imputed middle name) exactly matches the full name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 1% of patents are uniquely matched in this stage.
- **Stage 7**: Match to all information returns.
  - Name match: The inventor's full name exactly matches the full name of a taxpayer on any type of information return form.

- Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's information return forms.
- 6% of patents are uniquely matched in this stage.

B. A Comparison of the Firm Size Distribution in Census Data and EIN Size Distribution in Treasury Administrative Tax Files

Figure B.5.: Comparison of Census Firm Size and Treasury EIN Size Distributions, 2002



Notes: This figure shows the distribution of firm size in the Census distribution and EIN size in Treasury tax files, based on 2002 data. The distributions are very similar.

C. More Summary Statistics on Patent Inventor Teams

Table B.23.: Distribution of Outcomes for Two-Inventor Teams in 2002 (N=23,210)

		Oldest Team Member				
		1	2	3	4	5
	1	56	13	11	11	9
Age Quantile	2		16	31	30	21
of Youngest Team Member	3			21	43	34
	4				40	60
	5					100

		Richest Team Member				
		1	2	3	4	5
	1	24	26	24	12	14
Labor Earnings Quantile	2		26	44	16	14
of Poorest Team Member	3			36	35	28
	4				37	63
	5					100

		Richest Team Member				
		1	2	3	4	5
	1	17	33	23	9	17
Adjusted Gross Income Quantile	2		20	45	14	19
of Poorest Team Member	3			34	28	38
	4				26	73
	5					100

Notes: The numbers indicate the percentage of teams in each quantile bin, expressed as a share of all teams with their poorest/youngest team member in the same quantile. See Figure 1 in Section II of the paper for more details about the sample.

Figure B.6.: Frequency of Collaborations Across EINs

Team Size	N	Share w/ 1 EINs	Share w/ 2 EINs	Share w/ 3 EINs
2	262,198	0.73	0.27	-
3	148,100	0.65	0.26	0.08
4	73,636	0.59	0.27	0.10
5	33,496	0.53	0.28	0.12

Notes: This table shows the percentage of teams of various sizes collaborating across one or more EINs. For instance, the table reports that in 27% of two-inventor teams, the inventors are in two EINs, and that in 5% of five-inventor teams, the inventors are scattered across five EINs. Therefore, collabroations across EINs are quite frequent.

C. Appendix to Chapter 3

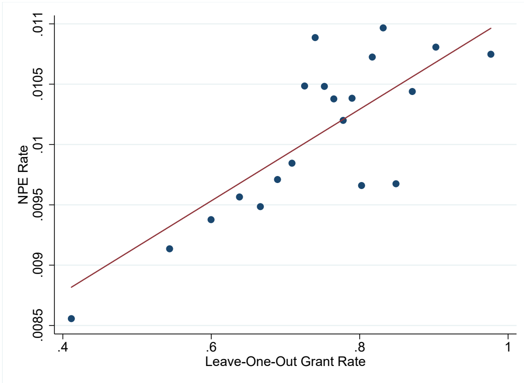
C.1. More Descriptive Statistics

Table C.1.: Correlation of Blocking Actions

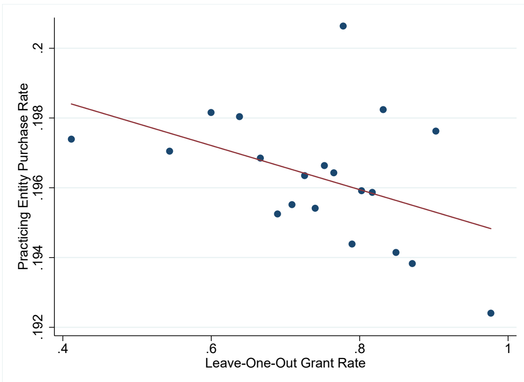
	101	102(a)	103(a)	112(b)
101	1.000	0.047	0.199	0.187
102(a)	-	1.000	0.104	0.086
103(a)	-	-	1.000	0.420
112(b)	-	-	-	1.000

Notes: Results are computed on the baseline sample (non-continuation granted patents covered by Frakes and Wasserman).

Figure C.1.: Preliminary Evidence of Examiner Impact



(a) All NPE purchases



(b) Non-NPE Purchase

Notes: Similar to Figure 3.1, but with examiner rates computed at the year by art unit level, and with art unit by year fixed effects.

Table C.2.: Art unit level statistics on the 670 art units in the PAIR data. Rates are unweighted, and are computed for units with more than 50 total cases in the period

Art unit level statistics	Median	Mean	Standard Deviation	Max
Examiners	13	17.9	19.7	201
Cases Processed	2552	3961	4190	23,164
Patents Granted	1536	2669	3100	17,252
NPE Patents	7	22.2	45.9	445
Grant Rate	0.67	0.65	0.16	1
NPE Patent Rate	0.006	0.009	0.010	0.057
Use of Section 101	0.040	0.093	0.108	0.450
Use of Section 102(a)	0.015	0.018	0.013	0.093
Use of Section 103(a)	0.42	0.42	0.11	0.78
Use of Section 112(b)	0.19	0.19	0.09	0.47

Table C.3.: Patents Asserted in Litigation (non-NPE)

Panel A: Primary Technology Categories

NBER ID	Category Name	Patents
6	Others	4,611
2	Computers & Communications	4,175
5	Mechanical	2,859
3	Drugs & Medical	2,609
4	Electrical & Electronic	2,497
1	Chemical	1,626
-	*New Classes (since 2001)*	708

Panel B: Secondary Technology Categories

NBER ID	Subcategory Name	Patents
22	Computer Hardware & Software	2,046
69	Miscellaneous-Others	1,729
21	Communications	1,614
31	Drugs	1,009
59	Miscellaneous-Mechanical	979
19	Miscellaneous-chemical	932
32	Surgery & Med Inst.	806
-	*New Classes (since 2001)*	708
51	Mat. Proc & Handling	620
33	Biotechnology	574
65	Furniture, House Fixtures	542
61	Agriculture, Husbandry, Food	504
55	Transportation	489
62	Amusement Devices	487
68	Receptacles	471
42	Electrical Lighting	451
49	Miscellaneous-Elec	442
41	Electrical Devices	409
45	Power Systems	394
46	Semiconductor Devices	364

Table C.4.: NPE Patent Holdings by NBER Technology Category (RPX data)

Panel A: Primary Technology Categories

NBER ID	Category Name	Patents
2	Computers & Communications	27,156
4	Electrical & Electronic	10,660
5	Mechanical	2,709
-	*New Classes (since 2001)*	2,324
1	Chemical	1,669
6	Others	1,453
3	Drugs & Medical	1,312

Panel B: Secondary Technology Categories

NBER ID	Subcategory Name	Patents
22	Computer Hardware & Software	11,459
21	Communications	11,020
46	Semiconductor Devices	4,667
24	Information Storage	3,298
-	*New Classes (since 2001)*	2,324
49	Miscellaneous-Elec	1,626
41	Electrical Devices	1,539
23	Computer Peripherials	1,379
19	Miscellaneous-chemical	1,159
45	Power Systems	1,121
54	Optics	906
59	Miscellaneous-Mechanical	816
42	Electrical Lighting	751
69	Miscellaneous-Others	745
32	Surgery & Med Inst.	716
43	Measuring & Testing	522
44	Nuclear & X-rays	434
39	Miscellaneous-Drgs&Med	370
51	Mat. Proc & Handling	360
55	Transportation	287

C.2. Random Assignment and Selection Effects

Random Assignment Falsification Tests

Table C.5.: Random Assignment Tests in Preferred Sample

Panel A: Patent's Predetermined Outcomes

	Patent's Number of Independent Claims	Number of Words in Patent's First Independent Claim
Leave-one-out Examiner Allowance Rate	0.0020177	0.0079458
	(0.0012327)	(0.0065522)
Class-artunit-year Fixed Effects	Yes	Yes

Panel B: Inventor's Predetermined Outcomes

	Inventor's Number of Patents Granted in Previous Years	Inventor's Number of Patents Litigated in Previous Years	Inventor's Number of NPE Patents in Previous Years
Leave-one-out Examiner Allowance Rate	0.0001682	$1.50 * 10^{-6}$	0.0002451
	(0.0001975)	$(1.19 * 10^{-6})$	0.0002188
Class-artunit-year Fixed Effects	Yes	Yes	Yes

Table C.6.: Random Assignment Tests in Full Sample

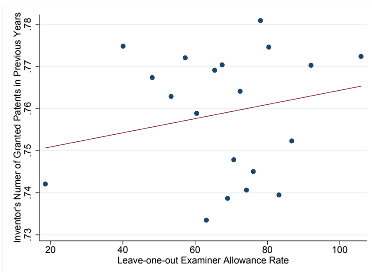
Panel A: Patent's Predetermined Outcomes

	Patent's Number of Independent Claims	Number of Words in Patent's First Independent Claim
Leave-one-out Examiner Allowance Rate	0.0064115***	0.0057961
	(0.0012471)	(0.0066375)
Class-artunit-year Fixed Effects	Yes	Yes

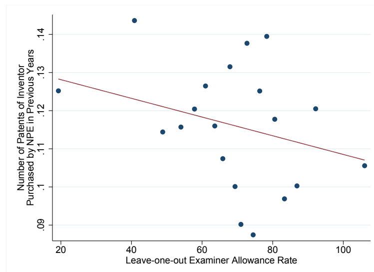
Panel B: Inventor's Predetermined Outcomes

	Inventor's Number of Patents Granted in Previous Years	Inventor's Number of Patents Litigated in Previous Years	Inventor's Number of NPE Patents in Previous Years
Leave-one-out Examiner Allowance Rate	0.0034767***	0.0000307***	0.0008802***
	(0.001084)	$(6.15 * 10^{-6})$	(0.0003454)
Class-artunit-year Fixed Effects	Yes	Yes	Yes

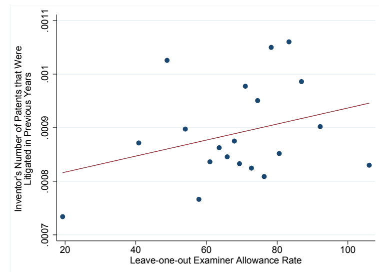
Figure C.2.: Random Assignment Tests in Preferred Sample



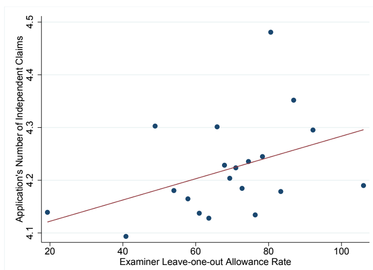
(a) Granted Patents



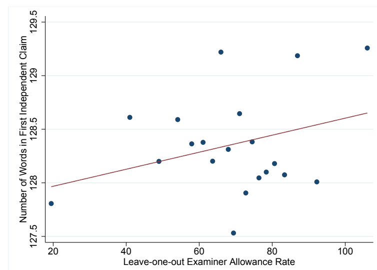
(b) NPE Patents



(c) Litigated Patents



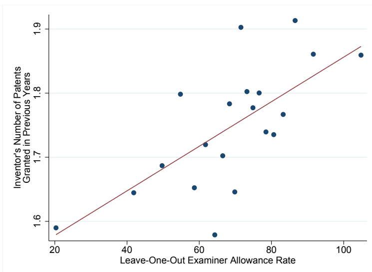
(d) Independent Claims



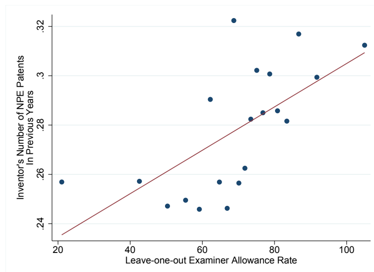
(e) Words in First Independent Claim

Notes: Sample excluding continuations and repeated inventor-examiner pairs. Leave-one-out examiner allowance rate adjusted for docket timing. Regressions include class-artunit-year fixed effects.

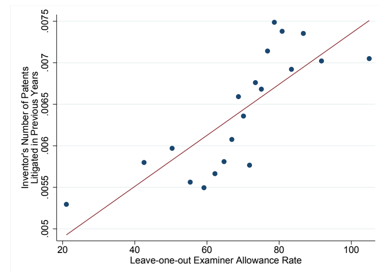
Figure C.3.: Random Assignment Tests in Full Sample



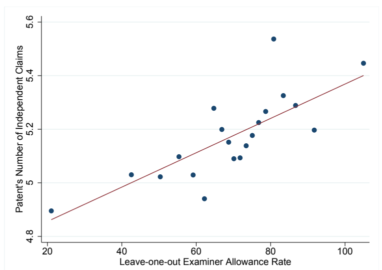
(a) Granted Patents



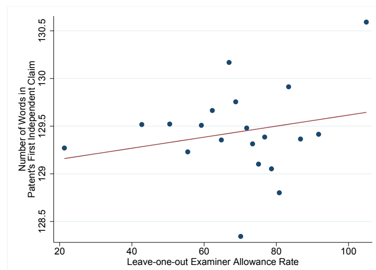
(b) NPE Patents



(c) Litigated Patents



(d) Independent Claims

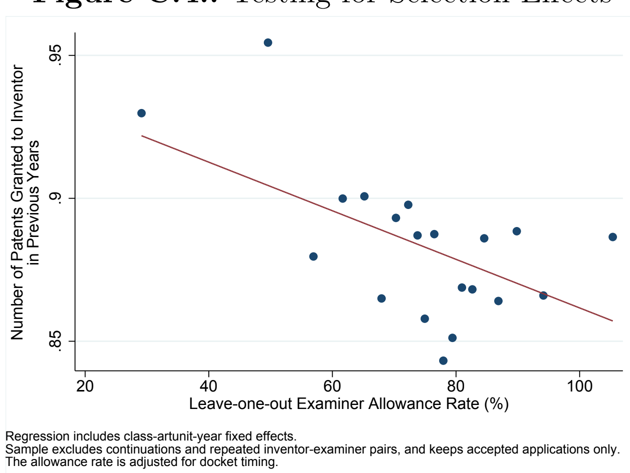


(e) Words in First Independent Claim

Notes: Regressions include class-artunit-year fixed effects.

Discussion of Selection Effects

Figure C.4.: Testing for Selection Effects



To understand the nature of selection effects in our setting and how we address them, consider the following framework. Within each artunit-year, each patent is defined by an examiner characteristic vector e (which includes examiner leniency, the propensity of the examiner to clarify the claims, etc...) and a patent "ex-ante" characteristic vector p (which can be thought of as the underlying technological quality of the patent). The nature of the intellectual property is the equilibrium outcome of the interaction between the examiner and the applicant and can therefore be thought of as a flexible function of e and p. We can therefore think of e and p as the primitives that span the intellectual property space (which makes explicit the notion that the examiner effectivel co-produces the intellectual property with the applicant).

We observe the following outcomes: whether a patent is granted (G), whether a patent is purchased by a regular entity (P), whether a patent is litigated by a regular entity (L), whether a patent is purchased by an NPE (P^{NPE}), whether a patent is litigated by an NPE (L^{NPE}), and whether a patent is invalidated in court I. In a partial equilibrium framework, these outcomes can all be thought of as a (possibly stochastic) function of the primitives e and p. Most of the outcomes of interest are subject to selection effects, which can be summarized as follows:

$$G = \begin{cases} 1 \; if \; g(e,p) > 0 \\ 0 \; else \end{cases}$$

$$P = \begin{cases} 1 \; if \; p(e,p) > 0 \; \& \; G = 1 \\ 0 \; else \end{cases}$$

$$L = \begin{cases} 1 \; if \; l(e,p) > 0 \; \& \; G = 1 \\ 0 \; else \end{cases}$$

$$P^{NPE} = \begin{cases} 1\ if\ p^{NPE}(e,p) > 0\ \&\ G = 1 \\ 0\ else \end{cases}$$

$$L^{NPE} = \begin{cases} 1\ if\ l^{NPE}(e,p) > 0\ \&\ G = 1 \\ 0\ else \end{cases}$$

$$I = \begin{cases} 1\ if\ i(e,p) > 0\ \&\ L = 1 \\ 0\ else \end{cases}$$

The definition above relate the osberved outcomes to unobserved latent variables of the form $f(e,p)$, which define mapping from the intellectual property space (spanned by e and p) to the outcomes of interest. For instance, an invalid patent is a patent for which $i(e,p) > 0$ - a well-defined mapping which can in theory be evaluated at any point of the intellectual property system. In practice, two challenges arise: first, we observe invalidation I only if the patent goes to litigation, which is itself an equilibrium outcome which depends on e and p; second, our empirical measures of e (examiner behavior) and p (underlying patent quality) are noisy.

Nonetheless, these definitions are useful to understand what we can learn from the data about the patent acquisition behavior of NPEs and, in particular, whether they tend to purchase "weak patents." Formally, we would like to test the following:

$$H_0 : \quad Cov(i(e,p), p^{NPE}(e,p)) > 0$$

where the covariance is computed over the intellectual property space. Testing whether NPEs purchase weak patents means testing whether they are active in parts of the intellectual property space that typically result in invalidations. The selection effects means that we only observe I and P^{NPE}, not the underlying mappings, and we need to find suitable measures of e and p in the data. In ongoing work, we relate our empirical findings with the theoretical covariance described above and clarify the assumptions under which the data allows us to conclude that NPEs tend to purchase weak patents.

C.3. Bayesian and Frequentist Interpretations of Shrinkage Methodology

Setup

Consider a setting with J examiners who are observed for T years. Assume that each examiner works in only one art unit and that each art unit has N students. The outcome of the granted patent is given by the following random effects model:

$A_{ijt} = \mu_j + \theta_{jt} + \epsilon_{it}$.

where i denotes a patent, j an examiner and t a year. A_{ijt} is the granted patent outcome (standardized by art unit-year), μ_j is an examiner effect (constant over time), θ_{jt} is an examiner-year shock, and ϵ_{it} is an idiosyncratic patent shock. We assume the following distributions: $\mu_j \sim N(0, \sigma_\mu^2)$, $\theta_{jt} \sim N(0, \sigma_\theta^2)$ and $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$. We are interested in estimating the vector of examiners' causal effects μ_j and using this vector for personnel policy.

Frequentist Approach

We want to forecast outcomes for patents granted by teacher j in year t using information on patent outcomes in all years prior to year t. Since all variables are demeaned, we don't need to include a constant in the set of predictors. Since there is no drift, we can use the average patent outcomes in all years prior to t, defined as $\bar{A}_{j,t-1} = \frac{1}{(t-1)N} \sum_{k=1}^{t-1} \sum_{i=1}^{N} A_{ijk}$, to predict patent outcome in year t. The optimal forecast minimizes the mean-squared error, so we must estimate α, defined in the population as:

$\alpha = argmin_b E[(A_{ijt} - a\bar{A}_{j,t-1})^2]$

In the random effect model above, providing an estimate $\hat{\alpha}$ of α amounts to providing estimates of examiner effects $\hat{\mu_j} = \hat{\alpha}\bar{A}_{j,t-1}$.

The first-order condition of the forecasting problem is:

$E[2(A_{ijt}\bar{A}_{j,t-1} - \alpha\bar{A}_{j,t-1})\bar{A}_{j,t-1}] = 0$

$\alpha = \frac{E[A_{ijt}\bar{A}_{j,t-1}]}{E[(\bar{A}_{j,t-1})^2]} = \frac{E[(\mu_j+\theta_{jt}+\epsilon_{it})\cdot(\mu_j+\frac{1}{t-1}\sum_{k=1}^{t-1}\theta_{jk}+\frac{1}{(t-1)N}\sum_{k=1}^{t-1}\sum_{i=1}^{N}\epsilon_{ik})]}{E[(\mu_j+\frac{1}{t-1}\sum_{k=1}^{t-1}\theta_{jk}+\frac{1}{(t-1)N}\sum_{k=1}^{t-1}\sum_{i=1}^{N}\epsilon_{ik})^2]}$

Using independence of examiner shocks across time and patent shocks across all patents within the same art unit - year,

$\alpha = \frac{E[\mu_j^2]}{E[\mu_j^2]+\frac{1}{t-1}E[\theta_{jk}^2]+\frac{1}{(t-1)N}E[\epsilon_{ik}^2]} = \frac{\sigma_\mu^2}{\sigma_\mu^2+\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)}$

In the sample, we can estimate α by estimating the variances σ_μ^2, σ_θ^2 and σ_ϵ^2 using the analogy principle, as described in the main text. This gives us $\hat{\alpha}$. So our estimator for μ_j is:

$\hat{\mu_j} = \frac{\hat{\sigma_\mu^2}}{\hat{\sigma_\mu^2}+\frac{1}{t-1}\left(\hat{\sigma_\theta^2}+\frac{\hat{\sigma_\epsilon^2}}{N}\right)}\bar{A}_{j,t-1}$

Bayesian Approach

We now derive an estimator of examiner effect in period t that is the posterior expectation of an examiner's effect given the history of patent outcomes up to period $t-1$.

Since there is no drift, we can use the average patent outcome in all years prior to t, $\bar{A}_{j,t-1}$ (defined above), as a sufficient statistic to form the posterior distribution of examiner effects. Using Bayes's rule, it is given by:

$\pi(\mu_j|\bar{A}_{j,t-1}) = \frac{f(\bar{A}_{j,t-1}|\mu_j)\phi(\frac{\mu_j}{\sigma_\mu})}{h(\bar{A}_{j,t-1})}$

where $f(\bar{A}_{j,t-1}|\mu_j)$ is the conditional probability density function (pdf) of $\bar{A}_{j,t-1}$ given μ_j, $\phi(.)$ is the pdf of the standard normal distribution (which we use given our prior forμ_j) and $h(\bar{A}_{j,t-1})$ is the unconditional pdf of $\bar{A}_{j,t-1}$. Since $\bar{A}_{j,t-1} = \mu_j+\frac{1}{t-1}\sum_{k=1}^{t-1}\theta_{jk}+\frac{1}{(t-1)N}\sum_{k=1}^{t-1}\sum_{i=1}^{N}\epsilon_{ik}$ and the shock are independent and normally distributed, we can write:

$$f(\bar{A}_{j,t-1}|\mu_j) = \phi\left(\frac{\bar{A}_{j,t-1}-\mu_j}{\sqrt{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)}}\right)$$

So we pick the vector of teacher effects in order to maximizes the posterior probability ofμ given the data:

$$max_{\{\mu_j\}_{j=1}^J}\ \Pi_{j=1}^J \pi(\mu_j|\bar{A}_{j,t-1}) \propto \Pi_j \left[\phi\left(\frac{\bar{A}_{j,t-1}-\mu_j}{\sqrt{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)}}\right)\phi(\frac{\mu_j}{\sigma_\mu})\right]$$

$$\propto \Pi_j exp(-\frac{(\bar{A}_{j,t-1}-\mu_j)^2}{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)} - \frac{\mu_j^2}{\sigma_\mu^2})$$

For each j, the first-order conditions with respect to μ_j give:

$$\frac{\bar{A}_{j,t-1}-\mu_j}{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)} - \frac{\mu_j}{\sigma_\mu^2} = 0$$

$$\frac{\bar{A}_{j,t-1}}{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)} = (\frac{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)+\sigma_\mu^2}{\sigma_\mu^2\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)})\mu_j$$

$$\mu_j = \frac{\sigma_\mu^2}{\frac{1}{t-1}\left(\sigma_\theta^2+\frac{\sigma_\epsilon^2}{N}\right)}\bar{A}_{j,t-1}$$

We can estimate the variances σ_μ^2, σ_θ^2 and σ_ϵ^2 using the analogy principle, as described in the main text, which yields an estimator that is identical to the frequentist estimator derived earlier:

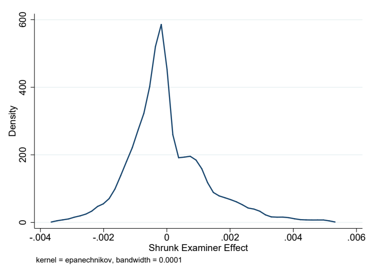
$$\hat{\mu_j} = \frac{\hat{\sigma_\mu^2}}{\hat{\sigma_\mu^2}+\frac{1}{t-1}\left(\hat{\sigma_\theta^2}+\frac{\hat{\sigma_\epsilon^2}}{N}\right)}\bar{A}_{j,t-1}$$

Intuitively, the Bayesian estimator is a precision-weighted average of the mean patent outcome for the patents granted by examiner j and the prior mean for μ_j, which is equal to 0. In other words, the mean test score observed in the data is shrunk towards 0 (empirical Bayes). As $\sigma_\theta^2 \to 0$ and $N \to \infty$, $\hat{\mu_j} \to \bar{A}_{j,t-1}$.

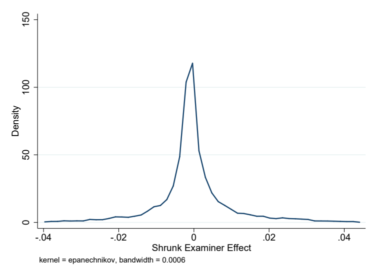
C.4. Additional Results on the Causal Effect of Patent Examiners

Causal Effect of Patent Examiners on Other Outcomes

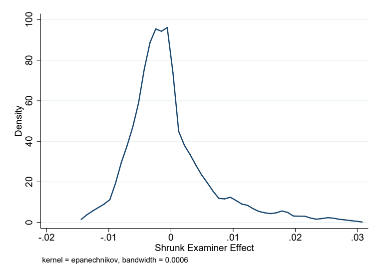
Figure C.5.: Distributions of examiner effects and tendencies by various outcomes observed in our dataset.



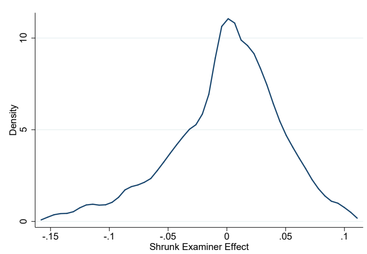
(a) Non-NPE Litigated Patents



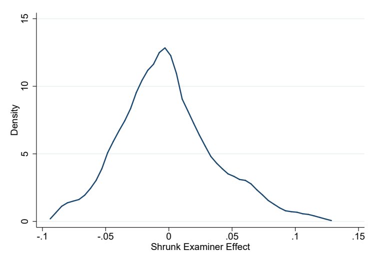
(b) 101 Rejections



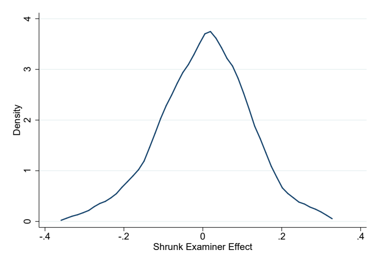
(c) 102(a) Rejections



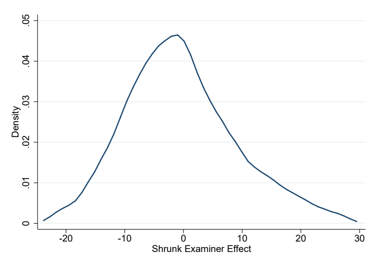
(d) 103(a) Rejections



(e) 112(b)



(f) Propensity to Add Independent Claims

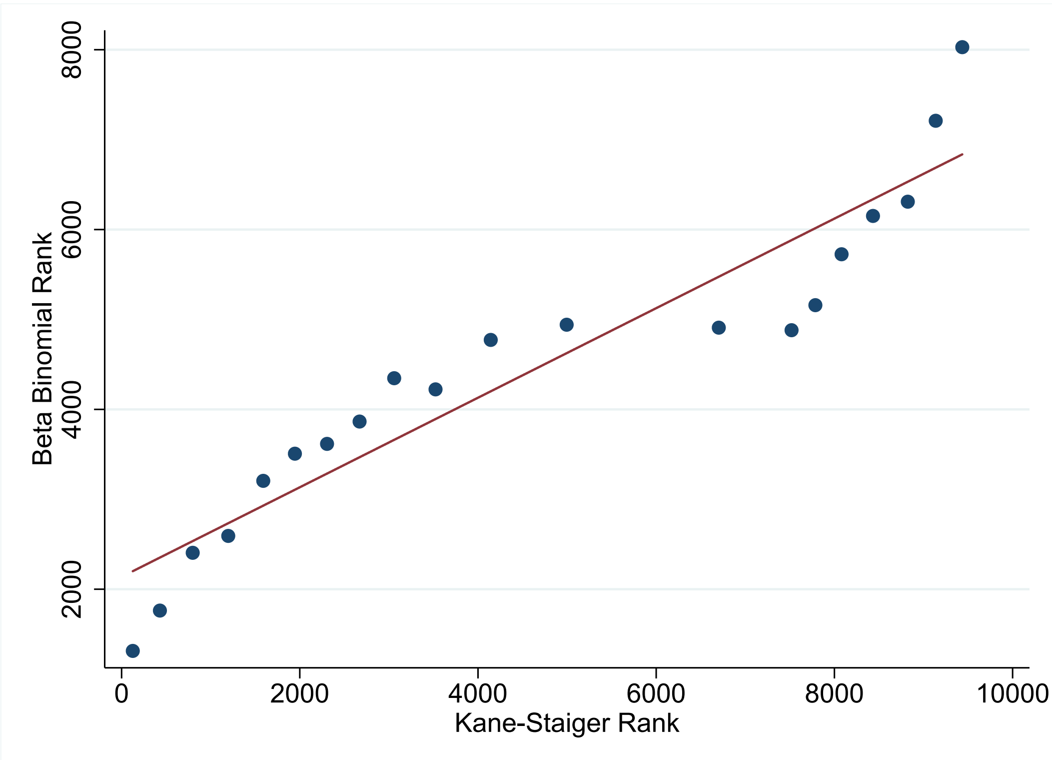


(g) Propensity to Add Words to Independent Claims

Notes: Distribution of examiner effects for other variables, computed in an analogous manner to Figure 3.2.

Comparison Between Kane-Staiger and Beta-Binomial Approaches

Figure C.6.: Comparison of Examiner Ranks



Notes: Comparison of examiner ranks, based on deviations computed through the Kane-Staiger methodology and through the Beta-Binomial methodology.

Blocking Action Robustness Check

In this section, we present additional robustness checks to the results shown in Section 3.5.2, by computing examiner blocking action causal effects using data from all applications they examine. This contrasts with our baseline specification, which only looks at blocking actions used on eventually granted patents. As shown in Table C.7, the results are generally very similar to the ones reported in Table 3.9.

Table C.7.: Pairwise Examiner Alternate Blocking Action Effects Results

	NPE Purchase	Non-NPE Purchase	NPE Litigated	non-NPE Litigated
101	-0.051*	0.003	-0.126*	-0.052**
	(0.022)	(0.004)	(0.064)	(0.018)
102(a)	0.013	0.003	-0.002	-0.011
	(0.023)	(0.005)	(0.051)	(0.018)
103(a)	-0.098***	-0.016**	-0.255***	-0.065**
	(0.024)	(0.005)	(0.064)	(0.020)
112(a)	-0.015	0.002	-0.128**	-0.054*
	(0.018)	(0.004)	(0.046)	(0.025)
112(b)	-0.047*	-0.003	-0.130*	-0.043*
	(0.021)	(0.005)	(0.057)	(0.019)
N	1,269,623	1,269,623	1,269,623	1,269,623

Notes: Similar to Tables 3.9 and 3.10, but computing examiner blocking action causal effects using their blocking actions on all applications. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

Signal Correlations

In this section, we present pairwise results using the signal correlation framework used in Chetty and Hendren (2015). The idea here is to split the sample of applications into two, and then compute the correlation of examiner effects across the samples. The signal correlation between examiner effects for variables X and Y is computed as follows:

$$\rho_{XY} = \frac{cov(E_{X1}, E_{Y2})}{\sqrt{cov(E_{X1}, E_{X2}) \cdot cov(E_{Y1}, E_{Y2})}}$$

The idea here is to look for correlation of examiner effects on different variables across samples, normalized by the consistency of the same-variable examiner effects across the samples. The method is similar to our previous pairwise predictive regressions, but incorporates reweighting of data points.

We report the results in Table C.8, performing a bootstrap routine at the examiner level to construct confidence intervals.

Table C.8.: Pairwise Examiner Alternate Blocking Action Effects Results

	NPE Purchase	Non-NPE Purchase	non-NPE Litigated
Word Count	-0.095*	0.011	-0.067*
	[-0.173,-0.041]	[-0.021,0.045]	[-0.151,-0.016]
101	-0.055*	0.043*	-0.054*
	[-0.121,-0.004]	[0.009, 0.076]	[-0.124,-0.011]
102(a)	0.028	0.027	-0.011
	[-0.017,0.082]	[-0.011,0.065]	[-0.063,0.040]
103(a)	-0.049*	0.037*	-0.028
	[-0.104,-0.008]	[0.008,0.067]	[-0.086,0.014]
112(a)	-0.002	0.047*	-0.054
	[-0.038, 0.032]	[0.013,0.083]	[-0.086,0.014]
112(b)	-0.027	0.026	-0.024
	[-0.073,0.011]	[-0.004, 0.060]	[-0.080,0.019]
N	1,269,623	1,269,623	1,269,623

Notes: 95 percent confidence intervals are shown in the brackets. ** p-value* < 0.05

Additional Legal Outcomes

In this section, we present results by predicting additional legal outcomes using examiner causal effects. The data is constructed from raw data provided by LexMachina.

Table C.9.: Pairwise Examiner Effect Results for Other Legal Outcomes

Panel A: District Court Litigation Outcomes

	Completed Litigation	Trial | Completed Lit.	Infringe | Trial	Invalid | Trial
NPE Purchase	0.050^{+}	0.015	-0.122**	0.103
	(0.026)	(0.036)	(0.045)	(0.088)
Words Per Claim Change	-0.195***	-0.010	0.057	0.044
	(0.018)	(0.040)	(0.054)	(0.076)
101	-0.068***	-0.009	-0.070	0.040
	(0.017)	(0.037)	(0.046)	(0.074)
102(a)	-0.001	0.033	-0.035	-0.134^{+}
	(0.019)	(0.045)	(0.058)	(0.081)
103(a)	-0.151***	-0.020	-0.010	0.036
	(0.021)	(0.040)	(0.057)	(0.080)
112(a)	-0.080***	0.088*	0.026	-0.009
	(0.017)	(0.04)	(0.049)	(0.071)
112(b)	-0.083***	-0.011	-0.059	-0.043
	(0.020)	(0.040)	(0.054)	(0.078)
N	1,867,760	5,319	595	595

Panel B: Inter-Partes Review Outcomes

	IPR Filed	Final Decision | Instituted	All Claims Invalid | Final Decision
NPE Purchase	0.116*	0.113*	0.015
	(0.055)	(0.054)	(0.042)
Words Per Claim Change	-0.246***	-0.014	-0.019
	(0.037)	(0.056)	(0.043)
101	-0.112**	-0.071	-0.036
	(0.040)	(0.054)	(0.034)
102(a)	0.015	0.032	-0.070
	(0.041)	(0.066)	(0.053)
103(a)	-0.167***	-0.022	-0.011
	(0.043)	(0.065)	(0.042)
112(a)	-0.103**	-0.058	0.012
	(0.034)	(0.052)	(0.031)
112(b)	-0.094*	-0.061	-0.019
	(0.039)	(0.057)	(0.041)
N	1,867,760	467	176

Notes: Similar to Table 3.9, but expanding the outcome dataset to incorporate all data from the PERD dataset, by assigning shrunk examiner effects to applications outside of the Frakes and Wasserman sample, while still using leave-one-out shrunk effects for applications within the Frakes and Wasserman sample. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

Application-Level Analysis

We present analogous results to the ones in Sections 3.4 and 3.5, but counting outcomes for rejected applications as zero. The results here capture both the intensive and extensive margin effects of an examiner. As we see in Table C.10, the extensive margin effects push many of the pairwise coefficients upwards in magnitude relative relative to the results reported in Table 3.9.

Table C.10.: Pairwise Examiner Blocking Action Effects Results (Application Level)

	NPE Purchase	Non-NPE Purchase	NPE Litigated	non-NPE Litigated
101	-0.071**	-0.018***	-0.143*	-0.071***
	(0.022)	(0.005)	(0.062)	(0.017)
102(a)	0.005	-0.005	-0.009	-0.018
	(0.024)	(0.006)	(0.052)	(0.017)
103(a)	-0.194***	-0.110*	-0.358***	-0.160***
	(0.026)	(0.006)	(0.068)	(0.020)
112(a)	-0.064***	-0.047***	-0.176***	-0.103***
	(0.019)	(0.006)	(0.046)	(0.025)
112(b)	-0.095***	-0.052***	-0.179**	-0.092***
	(0.023)	(0.006)	(0.057)	(0.019)
N	1,269,623	1,269,623	1,269,623	1,269,623

Notes: Similar to Tables 3.9 and 3.10, but computing examiner blocking action causal effects using their blocking actions on all applications. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

Instrumental Variable Analysis

As discussed earlier, the most commonly-used methodology in the examiner literature has been an instrumental variables approach, using the leave-one-out examiner grant rate as an instrument for allowance. This framework does not address the question we are interested in here, which is the effect of an examiner of the nature of the patent granted. Instrumental variables analysis is akin to performing the analysis in Section 3.4, without the extra analysis that we do in Section 3.5. In this section, we discuss the strong connection between the approaches, and also calculate the results using an instrumental variables approach.

Framework

The basic approach here is a standard instrumental variables analysis at the application level, instrumenting for approval using the leave-out-mean of examiner leniency. Formally, the first and second stage regressions are:

$$T_{ijat} = \beta_0 + \beta_1 Patented_{ijat} + Controls + \epsilon_{ijat}$$

$$Patented_{ijat} = \gamma_0 + \gamma_1 Z_{ijat} + Controls + \eta_{ijat}$$

where i denotes application, j denotes examiner, a denotes art unit, t denotes year of application. The reduced form regression looks like:

$$T_{ijat} = \alpha_0 + \alpha_1 Z_{ijat} + Controls + \zeta_{ijat} \tag{C.1}$$

The examiner leave-out-mean computed at the year-art unit level, akin to accounting for cohort effects, which in the teacher effects framework is captured by the θ_{jt} error term.

$$Z_{ijat} = \frac{n_{jat}^{grant} - Patented_{ijat}}{n_{jat}^{appl} - 1}$$

where n_{jat}^{grant} represents the number of granted patents for examiner j in art unit a in year t, and n_{jat}^{appl} represents the corresponding number for applications. More generally, we can instrument for examiner tendencies in a similar manner, replacing the *Patented* variable with an indicator for the usage of a given provision such as 103(a) and 112(b), and then instrumenting for the actual usage with a leave-one-out measure of average examiner usage of the given provision.

Comparison of Instrumental Variables to Kane-Staiger Framework

The instrumental variables (IV) framework has many similarities to the Kane-Staiger framework we applied in Sections 3.4 and 3.5.

The analysis in Section 3.5 is very similar in spirit to the reduced form IV specification (Equation C.1), as we leave out the current data point in computing examiner effects (a measure of examiner leniency in the IV setup), and then use it to predict an outcome T, such as NPE purchase. In the context of the IV-2SLS approach, the resulting coefficient α_1 is then scaled to by $\frac{1}{\gamma_1}$ in order to normalize the effect size, although in the case of the patent examiner instrument, γ_1 is often pretty close to 1. In addition, the motivation for computing Z at the year level rather than just at the examiner level is driven by the possibility of examiner by cohort effects. This is accounted for in our methodology through the θ_{jt} term.

There are a couple of major differences between the frameworks. First, the IV analysis is performed at the application level rather than the granted patent level. Therefore the reduced form coefficient α_1 will capture both the extensive margin (stricter examiners reject more applications, which then cannot be bought buy an NPE) and the intensive margin (stricter examiners also appear to force more edits to patent applications). Second, our mechanism analysis in Section 3.5 uses examiner effect measures computed across all years, rather than within a single year, reducing some of the noise in the measurement.

Results using IV Methodology

Our preferred specification is to include patent applications that were examined by examiners with 10 or more cases in a given year and art unit, which covers over 91% of the applications in our sample. We also run robustness checks with 5, 15 and 25 as the cutoff. Panel A of Table C.11 contains the first stage estimates. Consistent with the results reported in Sampat and Williams (2015), there is a very strong first stage relationship between the examiner leave-one-out grant rate and the decision on a given patent application. Some of the relationship is driven by year and art unit effects, but even after controlling for these, there is a precisely estimated coefficient of around 0.75.

Next, in Panel B of Table C.11, we report the coefficients from the two-stage least squares instrumental variables estimation. Here, we report the coefficients in a similar manner to the teacher effects results, by normalizing the coefficients to represent the effects of a one standard deviation change in examiner grant rate on NPE outcomes, normalized to the baseline NPE rate. Note that ex-ante, the reduced form relationship between NPE purchase and approval will be at least weakly monotonically increasing, because the NPE purchase vector is component-wise weakly less than the "Decision" variable. Due to the re-scaling, the OLS coefficient essentially captures the standard deviation in examiner grant rate, because the raw coefficient absent the scaling would just be the baseline rate of NPE purchasing, as rejected applications have a zero NPE purchase rate. The 2SLS coefficients capture the reduced form relationship between NPE purchase rates and examiner leave-one-out grant rate, scaled by the first stage coefficient. The fact that the 2SLS coefficient is higher than the OLS coefficient can be interpreted as evidence that the pool of NPE purchased patents are more "marginal," in the sense that they are more sensitive to examiner tendencies. A final point to note is that the coefficients are of similar magnitude to the ones found in the examiner effect distribution analysis. A one standard deviation change in examiner grant rate leads to an NPE effect of around 30% of the baseline effect.

Table C.11.: Instrumental Variables Results

Panel A: First Stage Results

Decision	(1)	(2)	(3)
Examiner Grant Rate	0.885***	0.752***	0.726***
	(0.00158)	(0.00185)	(0.00190)
Art Unit by Year FE		x	x
Class FE			x
N	2,478,697	2,478,697	2,478,697
R^2	0.180	0.186	0.189

Panel B: Examiner Impact on NPE Purchase

NPE Purchase	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS
Decision	0.284***	0.267***	0.387***	0.348***
	(0.00575)	(0.00549)	(0.0163)	(0.0184)
Art Unit by Year FE		x		x
N	2,478,697	2,478,697	2,478,697	2,478,697
R^2	0.003	0.018	0.003	0.002

Notes: Decision refers to the final disposal outcome of a given patent application. Examiner Grant Rate refers to the leave-one-out grant rate of an examiner in a given art uni and year. Analysis is restricted to examiners with more than 10 cases processed in a given year. Analysis is run on the entire Patent Examination Research Dataset. All standard errors are clustered at the examiner level. **p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.*

We also run similar analyses with the Frakes and Wasserman blocking action variables. The results are reported in Table C.12. The results are consistent with our earlier reported results. The usage of a given provision is strongly associated with usage of the provision on other applications (First Stage column). In addition, we see major discrepancies between the OLS and 2SLS results. This should not be entirely surprising, given our earlier summary statistics results (Table 3.5) and mechanism results. NPE purchased patents are almost equally like to receive 103(a) rejections and are more likely to receive 112(b) rejections (OLS results). However. the causal examiner contribution to this has the opposite sign, consistent with the result that examiners that tend to use more 112(b) blocking actions have fewer NPE purchased patents. As discussed earlier, the 2SLS coefficient also incorporates the direct effect of blocking actions on abandonments, so some applications mechanically cannot end up as NPE purchases.

This methodology could be useful for inferring the causal effects of specific blocking action types on outcomes, especially in cases where we can associated outcomes to ungranted patents, as in Sampat and Williams (2015) in relation to gene-based patent applications and subsequent research. However, the limitation is that if there is some correlation between usage of these provisions for a given examiner, and the instrumental variables results may pick up mechanisms associated with other provisions (it violates of the exclusion restriction). One could then think about using a more general instrumental variables framework with multiple predictors and multiple instruments, which would solve some of these problems as long as one assumes additive effects.

Table C.12.: Blocking Action Instrumental Variables Results

NPE Purchase	First Stage	OLS	2SLS
101	0.482***	-0.000547	-0.188***
	(0.00282)	(0.00498)	(0.0486)
102(a)	0.434***	0.00171	0.00137
	(0.00337)	(0.00222)	(0.0322)
103(a)	0.710***	-0.00919*	-0.217***
	(0.00248)	(0.00456)	(0.0297)
112(b)	0.690***	0.0103**	-0.130***
	(0.00252)	(0.00377)	(0.0247)
Art Unit by Year FE	x	x	x
N	1,752,641	1,752,641	1,752,641

Notes: Decision refers to the final disposal outcome of a given patent application. Examiner Grant Rate refers to the leave-one-out grant rate of an examiner in a given art uni and year. Analysis is restricted to examiners with more than 10 cases processed in a given year. Analysis is run on the Frakes and Wasserman coverage range. All standard errors are clustered at the examiner level. **p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.*

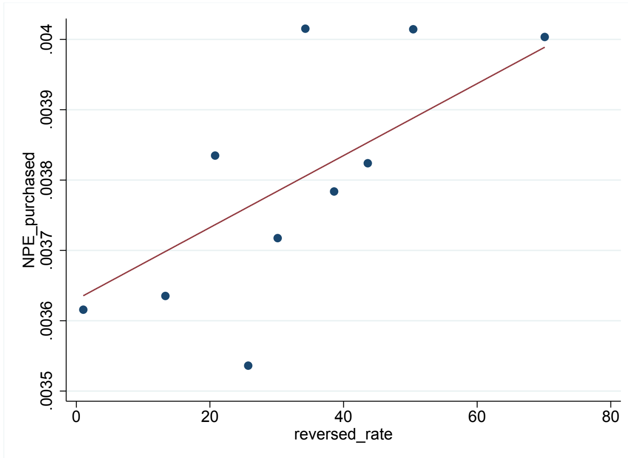
C.5. Additional Results on Mechanisms

Table C.13.: Results across technology centers for key predictors of NPE purchase and non-NPE litigation.

	NPE Purchase		Non-NPE Litigation	
Specification	103(a)	112(b)	103(a)	112(b)
1600 - Biotechnology and Organic Chemistry	-0.173	0.078	-0.036	0.002
	(0.120)	(0.077)	(0.068)	(0.067)
1700 - Chemical and Materials Engineering	0.072	0.187	0.025	0.052
	(0.130)	(0.173)	(0.037)	(0.035)
2100 - Computer Architecture, Software, and Information Security	-0.064**	-0.036	-0.124**	-0.121**
	(0.023)	(0.026)	(0.005)	(0.038)
2400 - Computer Networks, Multiplex communication, Video Distribution, and Security	-0.016	0.024	-0.119^{+}	0.003
	(0.029)	(0.030)	(0.062)	(0.050)
2600 - Communications	-0.098***	-0.061**	-0.010	-0.070*
	(0.025)	(0.023)	(0.030)	(0.031)
2800 - Semiconductors, Electrical and Optical Systems and Components	-0.062**	-0.036^{+}	-0.071**	-0.013
	(0.022)	(0.021)	(0.024)	(0.024)
3600 - Transportation,Construction, Electronic Commerce, Agriculture, National Security...	-0.049	-0.062	-0.070*	-0.069*
	(0.039)	(0.039)	(0.029)	(0.027)
3700 - Mechanical Engineering, Manufacturing, Products	0.034	-0.046	-0.001	-0.002
	(0.040)	(0.043)	(0.029)	(0.028)

Notes: Same specifications as Tables 3.9 and 3.10, but restricting the sample to each technology center. Baseline NPE purchase and non-NPE litigation rates are re-computed for each technology center. *+ p-value <0.10, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001*

Figure C.7.: NPEs Tend to Purchase Patents from Examiners Who Are Reversed at PTAB



C.6. Data Appendix

Name Matching

In this part, we discuss our procedure for creating NPE portfolios. The algorithm proceeds as follows:

1. We start off with a list of NPE names, either from RPX or from Cotropia et al.
2. We normalize entity names from both the NPE list and the USPTO Assignment Database. This is done by capitalizing all names, removing punctuation, and removing standard entity terms: INC, CO, COMPANY, COMPANIES, CORP, CORPORATIONS, DIV, GMBH, LLC, LC, INCORPORATED, KG, LIMITED, LIMITED PARTNERSHIP, LP, LTD, NV, PLC, SA, SARL, SNC, SPA, SRL, TRUST USA, CENTER, BV, AG, AB, GROUP, FOUNDATION, INSTITUTE, and TECHNOLOGIES.
3. We then collect Reel/Frame IDs of patent transactions in the USPTO Assignment Database that have a normalized entity name matching the normalized name of an NPE

4. Each Reel/Frame ID is classified in the USPTO data based on the type of transaction and whether the assignment was to an employer (essentially the first assignment). We keep transactions that are non-employer assignments. This then gives us a set of patents involved in patent purchasing, and excludes other types of transactions such as securitization, mergers, and name changes.
5. Finally, we collect the list of patents associated with each of these transactions to create our portfolios.

D. Appendix to Chapter 4

D.1. Appendix Figures and Tables

Table D.1.: Additional Summary Statistics

Table A1 Panel A:
Summary Statistics for Inventors Matched to Parents - all

	Mean	Median	Standard Deviation
Patent Grants	3.30	1	9.07
Lifetime Citations	28.54	2	154.57
Age at Patent	44.72	43	28.36
Wage Income ($)	112,117	81,000	252,276
Total Income ($)	178,710	109,000	475,458
Female	11.6%	0	

Number of Inventors: 1,200,620

Table A1 Panel B:
Summary Statistics - 1980-84 Birth Cohorts

	Mean	Median	Standard Deviation
Patent Grants	1.37	1	3.21
Patent Applications	2.04	1	3.94
Lifetime Citations	1.16	0	15.04
Age at Patent	27.45	27	2.43
Wage Income ($)	27,296	8,000	48,000
Total Income ($)	37,270	10,000	131,298

Number of Inventors: 45,083

Table D.2.: Regression Estimates for Relationship between Math and Verbal Test Scores

Table A2 : Patent Rates vs. Math and Verbal Test Scores: Regression Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:			**Number of inventors (per 1,000)**				**Number in top 1% of Adult income**	
3rd Grade Math Score	0.85*** (0.06)		0.84*** (0.10)	0.91*** (0.10)		8.5*** (0.26)		6.6*** (0.35)
3rd grade English Score		0.68*** (0.06)	0.05 (0.10)		0.08 (0.09)		7.7*** (0.26)	2.8*** (0.32)
Math Vingtile FE					X			
English Vingtile FE				X				
Observations	223,100	214,900	210,328	210,328	210,328	247,537	238,077	232,874

Notes: *** significant at 1% level, **5% level, *10% level. OLS estimates with standard errors in parentheses below coefficients. NYC data. Test scores measures are standardized to have mean 0 and standard deviation 1. "Math (English) vingtile FE" are 20 dummy variables/fixed Effects for each Math (English) test score vingtile. Dependent variable in columns (1) to (5) is whether child becomes inventor by age 32. Dependent variable in Columns (6)-(8) is whether child ends up in the top 1% of the income distribution for her cohort by age 32.

Table D.3.: Percent of Innovation Gap Accounted for by Test Scores - Alternative Estimators

Table A3: Percent of Gap Accounted for by Test Scores - Alternative Estimators

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline DFL	**Balanced Panel DFL**	**Median Split DFL**	**Baseline FE**	**Balanced Panel FE**	**Median Split FE**
Grade 3	30.9%	35.5%	27.8%	39.4%	41.5%	43.0%
Grade 4	36.4%	42.9%	32.7%	42.0%	44.7%	48.3%
Grade 5	39.2%	45.8%	42.5%	47.4%	48.2%	57.7%
Grade 6	44.8%	46.3%	42.4%	48.8%	47.8%	52.8%
Grade 7	50.5%	55.1%	49.9%	55.1%	52.2%	59.0%
Grade 8	52.0%	50.1%	49.3%	56.5%	51.5%	61.0%

Notes: These are experiments to assess the fraction of the relationship between parental income and whether a child grows up to be an inventor. The data is the same as underlying Table 1 and column (1) reproduces the results in column (2) of that Table. Column (2) uses the sub-sample Of individuals for whom we observe all years of test score data grades 3-8. Column (3) uses a split at the median of parent income instead of At the 80th percentile. Columns (4)-(6) reproduce the specifications of the first three columns except that instead of using a DiNardo et al (1996) Approach we simply add in 20 vingtile fixed effects and a dummy for being above the 80th percentile of the income distribution (or median in Final column).

Table D.4.: Illustration of Technology Class and Distance

Table A4: Illustration of Technology Classes and Distance

Category: ***Computers + Communications***

Sub-category: ***Communications***

Technology Class (= 375)	**Distance Rank**
Pulse or digital communications	0
Demodulators	1
Modulators	2
Coded data generation or conversion	3
Electrical computers: arithmetic processing and calculating	4
Oscillators	5
Multiplex communications	6
Telecommunications	7
Amplifiers	8
Motion video signal processing for recording or reproducing	9
Directive radio wave systems and devices (e.g., radar, radio navigation)	10

Table D.5.: Patent Rates by Commuting Zone Where Child Grew Up

Table A5 Patent Rates by Commuting Zone (CZ) Where Child Grew Up
Top 10 and Bottom 10 CZs Among 100 Largest CZs

Top 10 CZs			**Bottom 10 CZs**		
Rank	CZ	Inventors per Thousand	Rank	CZ	Inventors per Thousand
1	San Jose, CA	5.41	91	Birmingham, AL	1.03
2	Madison, WI	4.87	92	Virginia Beach, VA	1.01
3	Minneapolis, MN	4.29	93	El Paso, TX	0.92
4	San Francisco, CA	3.83	94	Fresno, CA	0.88
5	Detroit, MI	3.78	95	Little Rock, AR	0.83
6	Boston, MA	3.75	96	Modesto, CA	0.82
7	Allentown, PA	3.65	97	Fayetteville, NC	0.82
8	Milwaukee, WI	3.54	98	Lakeland, FL	0.79
9	Manchester, NH	3.50	99	Mobile, AL	0.69
10	Albany, NY	3.27	100	Brownsville, TX	0.60

Table D.6.: Percent of Female Inventors Across CZs

Table A6 Percent of Inventors who are Female: Top 10 and Bottom 10 CZs
Among the 100 CZs with the Largest Number of Inventors

Top 10 CZs			Bottom 10 CZs		
Rank	CZ	Percent Female	Rank	CZ	Percent Female
1	Toms River, NJ	28.4	91	Eau Claire, WI	12.7
2	Jacksonville, FL	26.3	92	Salt Lake City, UT	12.4
3	Dayton, OH	23.9	93	Rochester, MN	12.0
4	Charlotte, NC	23.8	94	Erie, PA	11.9
5	Louisville, KY	23.7	95	Peoria, IL	10.9
6	Atlanta, GA	23.7	96	Fort Wayne, IN	10.8
7	Portland, ME	23.2	97	Fort Collins, CO	10.1
8	Miami, FL	23.2	98	Fresno, CA	9.7
9	Raleigh, NC	22.8	99	Oklahoma City, OK	9.5
10	Poughkeepsie, NY	22.7	100	Santa Rosa, CA	8.9

Figure D.1.: Patent Rates between Ages 30-40 vs. Parent Income Percentile

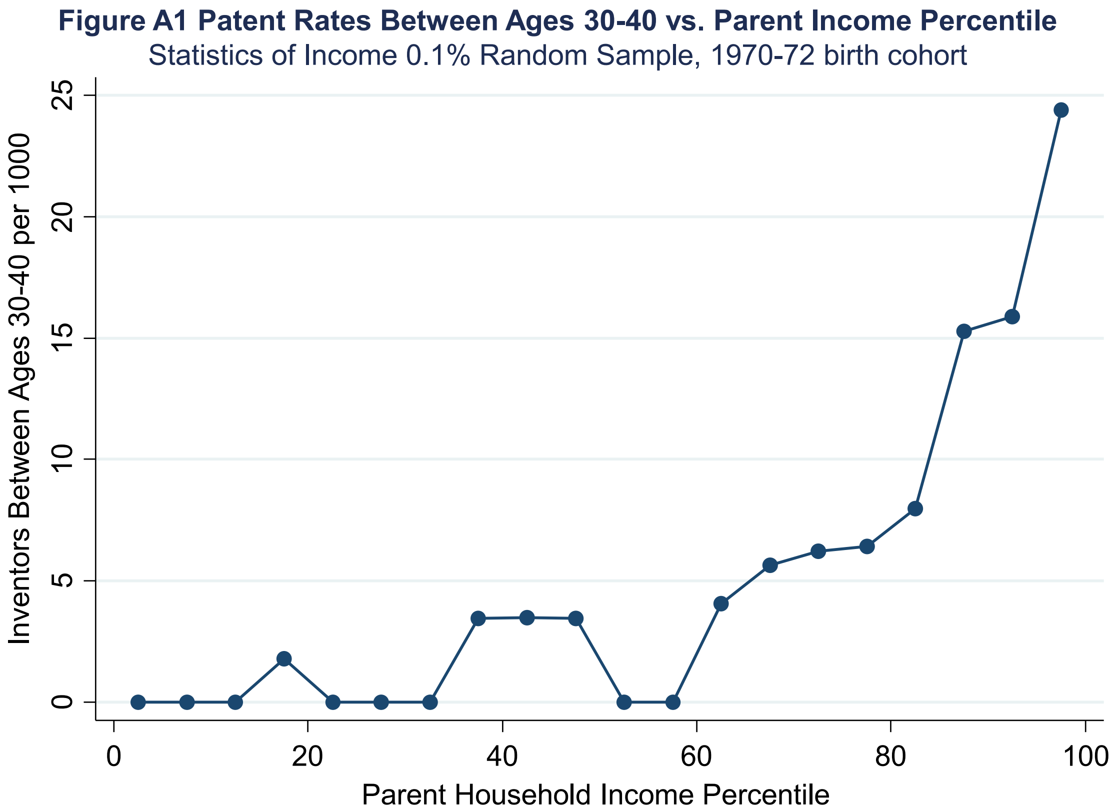


Figure D.2.: Patent Rates vs. Parent Income Percentile

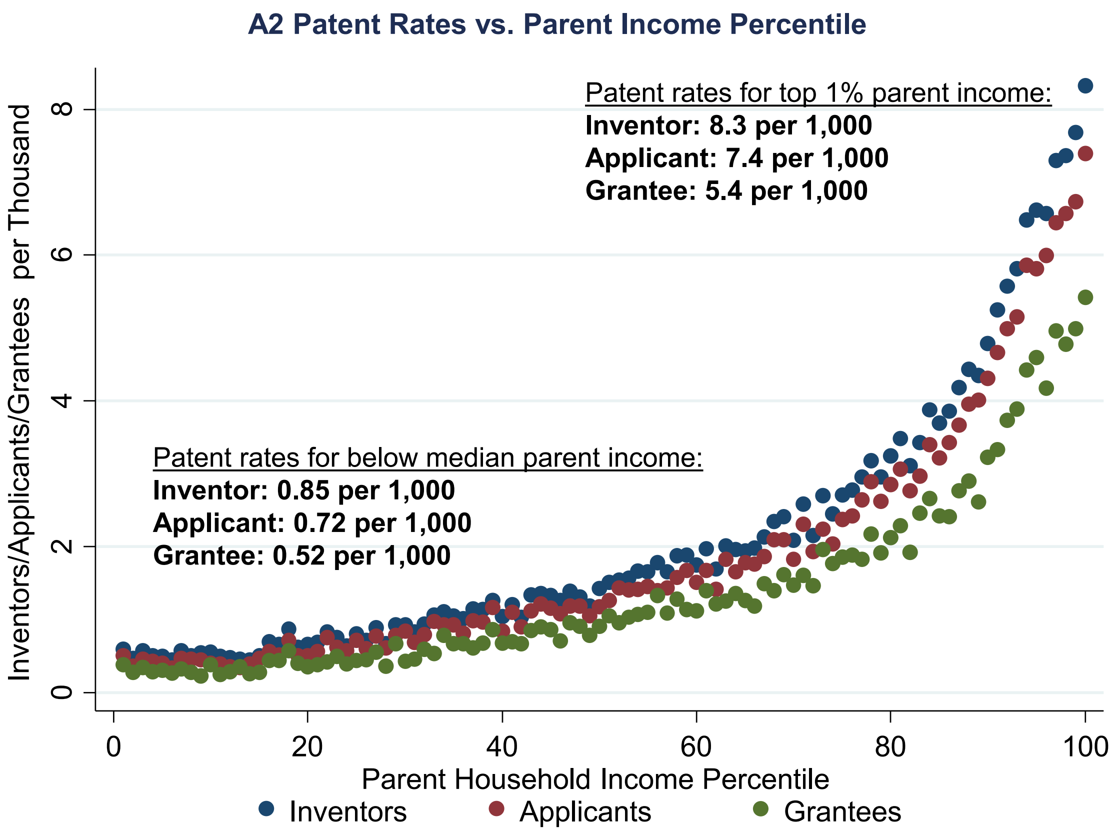


Figure D.3.: Percentage of Children in Top 1% of Cohort's Income Distribution vs. Parent Income Percentile

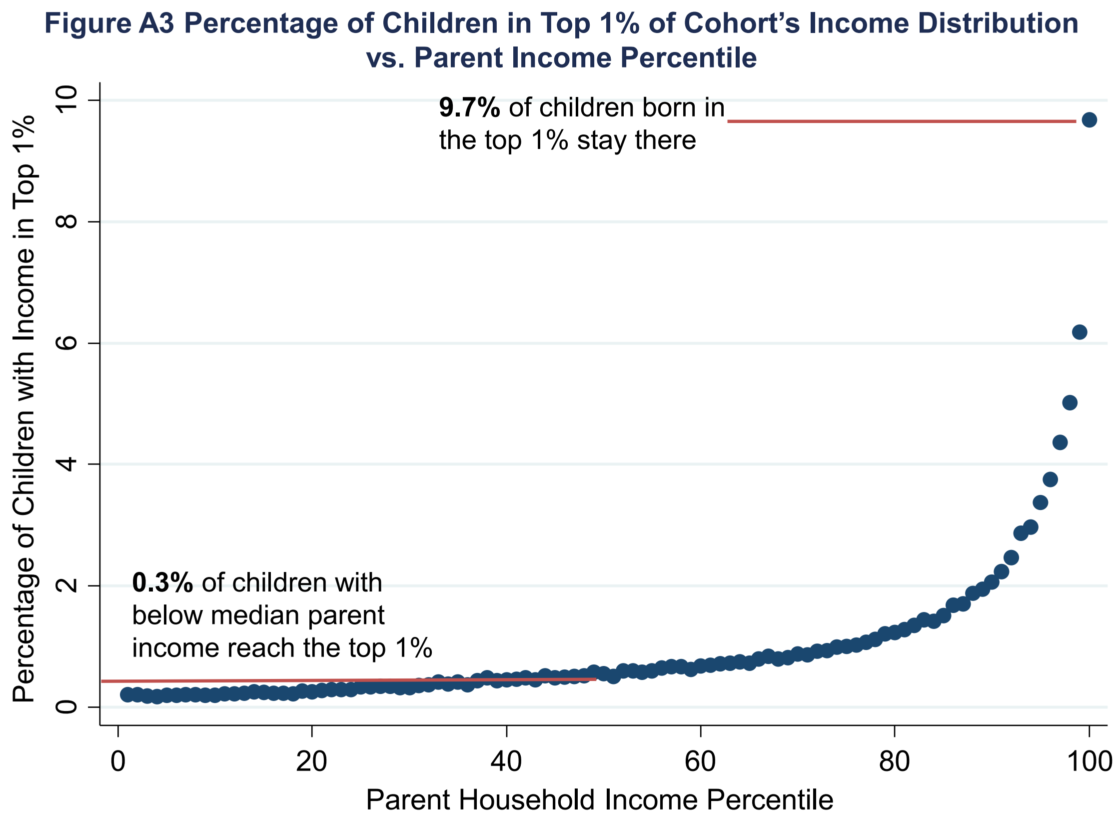


Figure D.4.: Patent Rates vs. Parent Income in NYC Public Schools

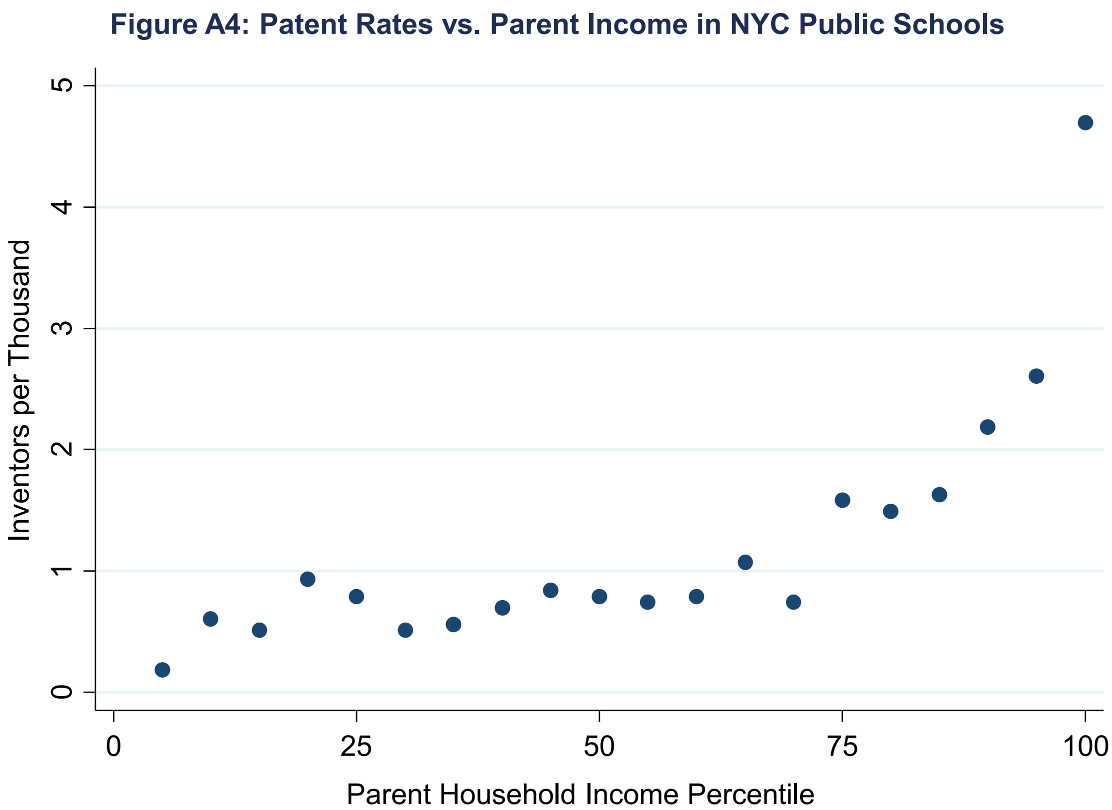


Figure D.5.: Concentration of Innovation Among Graduates of Selected Colleges

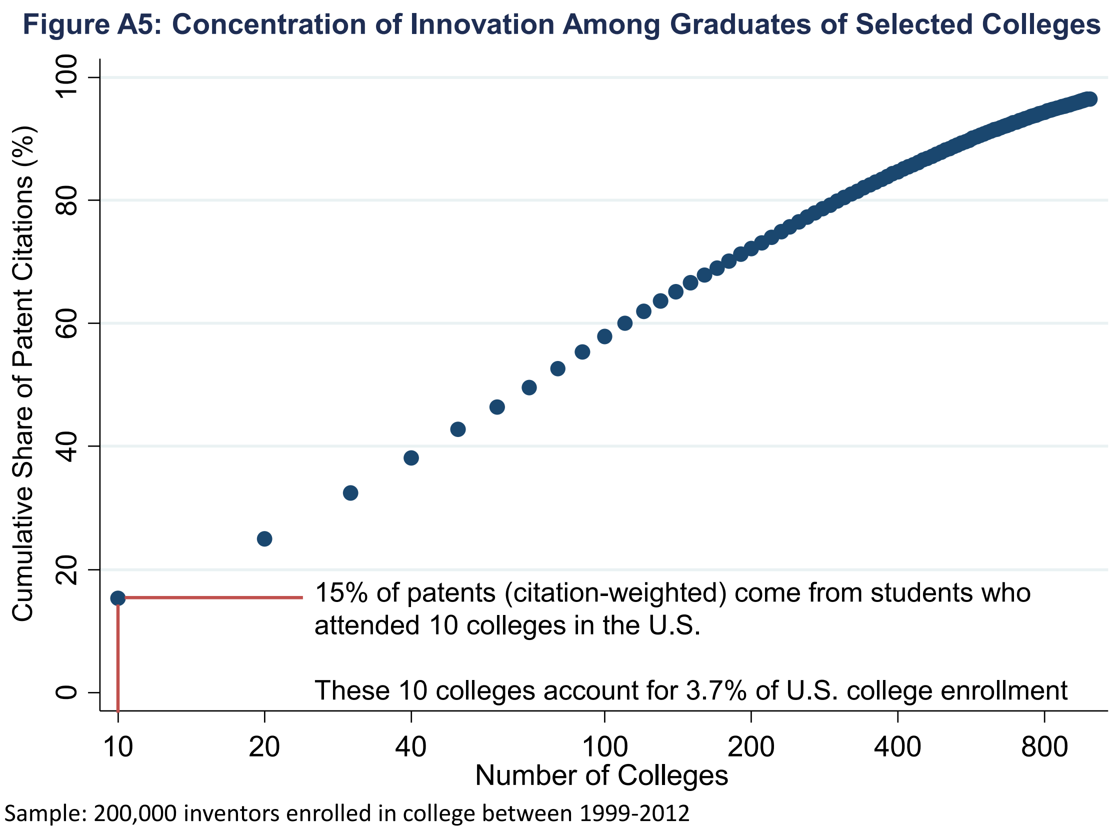


Figure D.6.: Innovation at Age 30 Rates by College

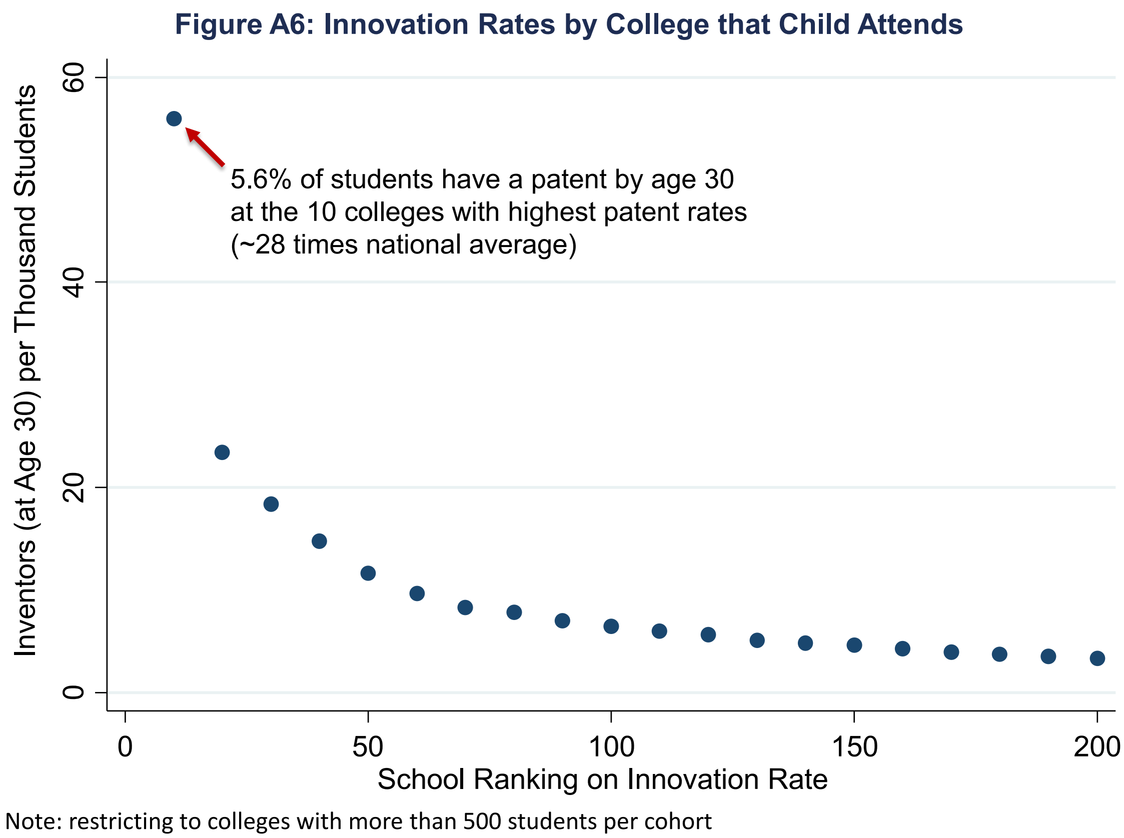


Figure D.7.: Innovativeness of College vs. Parent Income Percentile

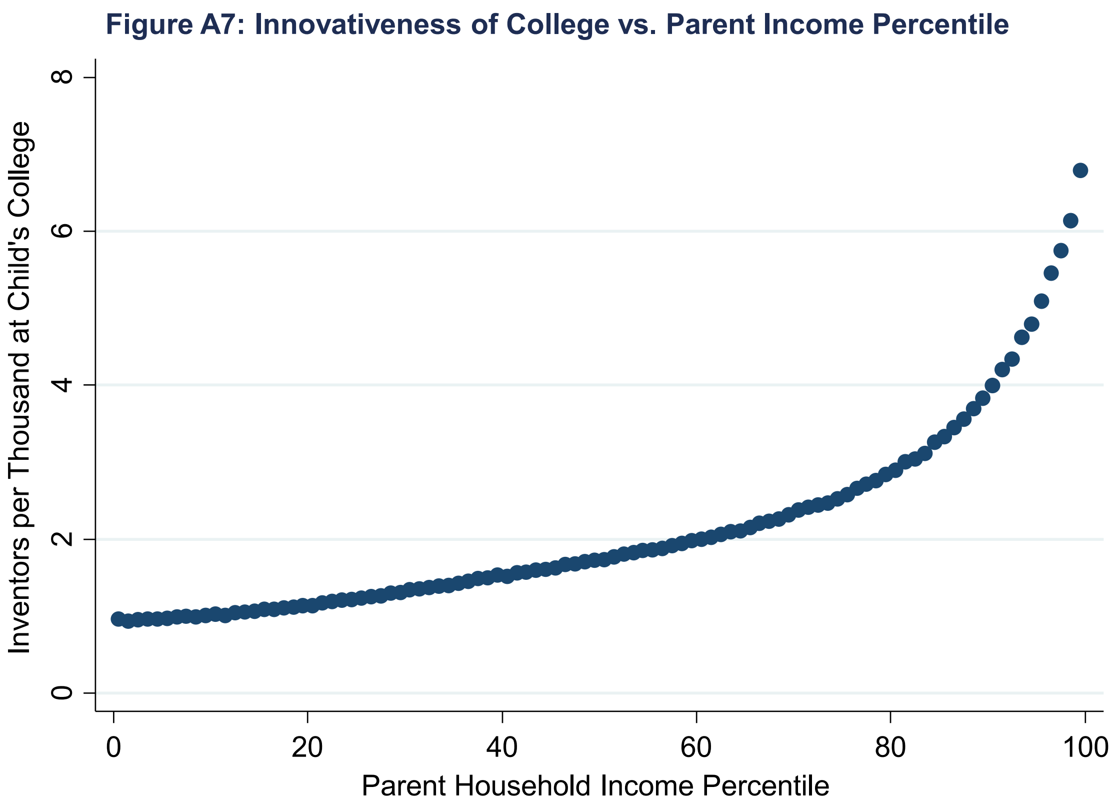
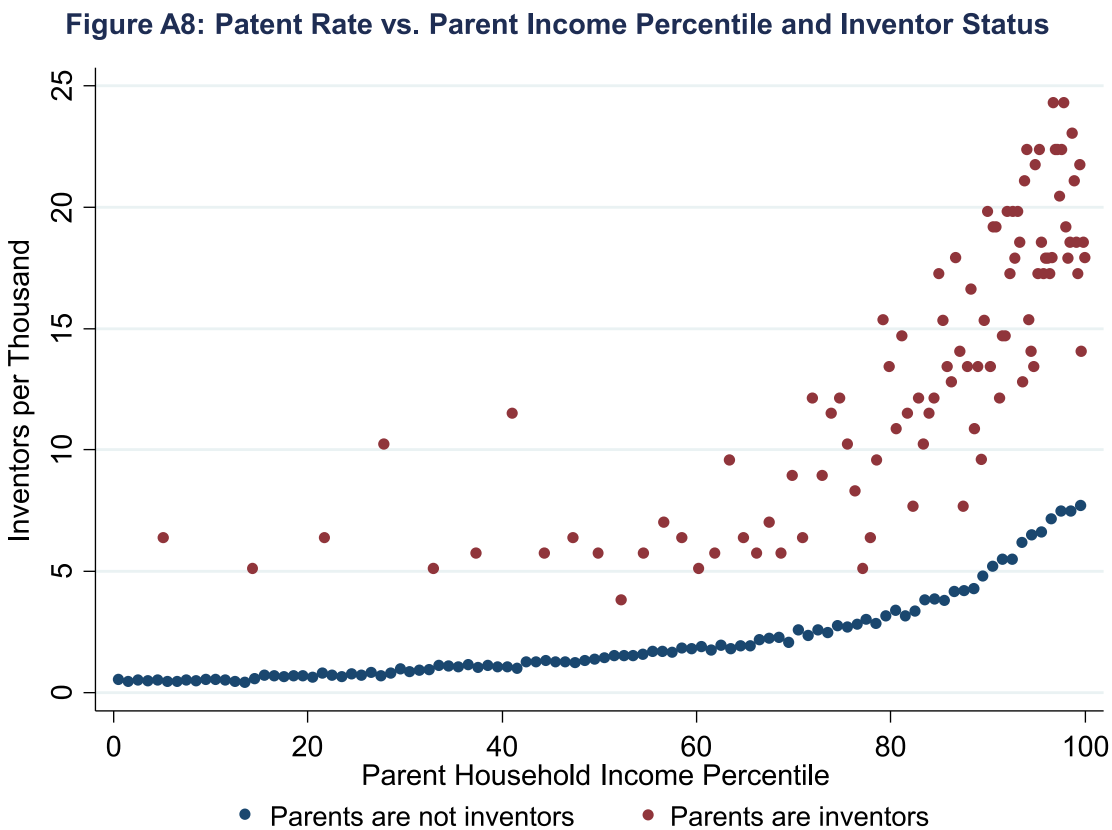


Figure D.8.: Patent Rate vs. Parent Income Percentile and Inventor Status



D.2. Data Appendix

A.1 Data Preparation

- **Suffix Standardization**. Suffixes may appear at the end of taxpayers' first, middle, or last name fields. Any time any of these fields ends with a space followed by "JR", "SR", or a numeral I-IV, the suffix is stripped out and stored separately from the name[1].

- **First name to imputed first/middle name**. The USPTO separates inventor names into "first" and "last," but the IRS often separates names into first, middle, and last. In practice, many inventors do include a middle initial or name in the first name field. Whenever there is a single space in the inventor's first name field, for the purposes of matching, we allow the first string to be an imputed first name, and the second string to be an imputed middle name or initial. The use of these imputed names is outlined below.

A.2 Pseudocode for Match on Name and Location

The exact matching stages are as follows. We conduct seven progressive rounds of matching. Inventors enter a match round only if they have not already been matched to a taxpayer in an earlier round. Each round consists of a name criterion and a location criterion. The share of data matched in each round is noted, with an impressive 49% being exact matches on the first stage.

- The matching algorithm takes as input a relation of inventor data and five relations of IRS data:

  - Input relations:

[1] Numerals I and V are only permissive suffixes at the end of a last name field, as these may be middle initials in a middle name field.

    * Inventors(inv_id, first, last, imputed_first, imputed_middle, suffix) - directly from USPTO
    * NamesW2(irs_id, first, middle, last, suffix) - all names used by individual on W2 information returns; name field is recorded as first, middle, and last
    * Names1040(irs_id, first, middle, last) - all self-reported names from 1040 forms[2]
    * Nameln1W2(irs_id, fullname) - all names from W2, but a separate variable not recorded as first, middle, last that was more frequently present
    * CitiesW2(irs_id, city, state) - all cities reported on W2
    * Zips1040(irs_id, name) - all zip codes reported on 1040
  - Output relation:
    * Unique-Matches (inv_id, irs_id)

- **Stage 1**: Exact match on name and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. Either the inventor's first name field exactly matches the concatenation of the IRS first and middle name fields or the IRS middle name field is missing, but the first name fields match. If an imputed middle name is available for the inventor, candidate matches are removed if they have ever filed at the IRS with a middle name or initial that conflicts with the inventor's.
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 49% of patents are uniquely matched in this stage.

[2]We only take names off of 1040s for those who file singly because it proved difficult to parse names of those list them jointly

- **Stage 2**: Exact match on imputed name data and location.
  - Name match: The inventor's last name exactly matches the taxpayer's last name and the taxpayer's last name is the same as the inventor's imputed first name. Either the inventor's imputed middle name/initial matches one of the taxpayer's middle/initial name fields, or one of the two is missing. For inventors with non-missing imputed middle names, priority is given to matches to correct taxpayer middle names rather than to taxpayers with missing middle names. As above, candidate matches are removed if they have ever filed at the IRS with a conflicting middle name or initial.
  - Location match: As above, the inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 12% of patents are uniquely matched in this stage.
- **Stage 3**: Exact match on actual or imputed name data and 1040 zip crosswalked.
  - Name match: The inventor's last name exactly matches the taxpayer's last name. The inventor's first name matches the taxpayer's first name in one of the following situations, in order of priority:

  1. Inventor's firstname is the same as the taxpayer's combined first and middle name.
  2. Inventor's imputed firstname matches taxpayer's and middle names match on initials.
  3. The inventor has no middlename data, but inventor's firstname is the same as the taxpayer's middle name.

  - As always, taxpayers are removed if they are ever observed filing with middle names in conflict with the inventor's.Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.

  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 3% of patents are uniquely matched in this stage.

- **Stage 4**: Same as previous stage, but using 1040 names instead of names from W2's.
  - Name match: The inventor's name matches the name of a 1040 (or matches without inventor's middle initial/name, and no taxpayer middle initials/names conflict with inventor's).
  - Location match: The inventor's city and state must match some city and state reported by that taxpayer exactly.
  - 6% of patents are uniquely matched in this stage.

- **Stage 5**: Match using W2 full name field.
  - Name match: The inventor's FULL name exactly matches the FULL name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.
  - 8% of patents are uniquely matched in this stage.

- **Stage 6**: Relaxed match using W2 full name field.
  - Name match: The inventor's full name (minus the imputed middle name) exactly matches the full name of a taxpayer on a W2.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's 1040 zip codes.

- 1% of patents are uniquely matched in this stage.

- **Stage 7**: Match to all information returns.
  - Name match: The inventor's full name exactly matches the full name of a taxpayer on any type of information return form.
  - Location match: The inventor's city and state match one of the city/state fields associated with one of the taxpayer's information return forms.
  - 6% of patents are uniquely matched in this stage.

A.3 Final matched sample

As noted in the text we matched 88% of inventors in the 200s and 80% in the 1990s. The match seemed balanced by characteristics of the patents such as citation rates, number of claims, technology class, zip code, etc. The match rate is slightly lower among self-assigned patents.

D.3. Lifecycle Model of Innovation

B.1 Basic Set up

B.1.1 Model

We sketch a simple inventor lifecycle model. Consider a two period model. In period 1, individual i begins with an endowment of human capital we call "ability" (a_i). Their human capital then evolves over the course of their schooling increasing by s_i which is determined by family, school and neighborhood quality. Initial ability is complementary with future learning and parental income is positively associated with both initial ability and subsequent inputs. We initially assume these are exogenous, but later allow s_i to be endogenous. After acquiring human capital, $H_i = H(a_i, s_i)$, individuals enter the labor market in period 2 and choose an occupation. For now we consider just two possible occupations: the R&D sector and the Non-R&D sector.

Wages in the non-R&D sector are $\widetilde{w} + \rho H_i$.[3] The expected utility (V^N) for working in the non-R&D sector will depend on idiosyncratic tastes for working in this sector which we denote in utility terms as v_i^N. There is also a tax schedule *T(.)* which we will allow to be non-linear with a higher marginal rate above an upper threshold. So $V_i^N = u[T(\widetilde{w} + \rho H_i)] + v_i^N$ where *u[.]* is a utility function over consumption.

In the R&D sector workers also receive a deterministic base wage ρH_i but also have a chance of receiving an additional stochastic reward from innovation. We assume that those with more human capital have a higher chance of successfully innovating, so the additional (potential) innovation reward is $\pi(H_i)$, with $\pi'(H_i) > 0$. For simplicity we parameterize this as $\pi(H_i) = \pi H_i$, We allow a taste term for the R&D sector v_i^R which implies that the expected value of choosing to work in the R&D sector is:

$$V_i^R = E_\pi(u[T(\rho H_i + \pi H_i)]) + v_i^R$$

where $E_\pi(.)$ is the expectations operator taken over the stochastic innovation variable π.

We now introduce the idea that some individuals who are less exposed to inventors underestimate the real returns to invention. The *perceived* value of a career in the R&D sector is:

$$V_i^{PR} = \lambda_i V_i^R + (1 - \lambda_i) E_\pi(u[T(\rho H_i + \eta)]) + v_i^R$$

where $0 \leq \lambda \leq 1$ and $\eta < \pi H_i$. With probability λ an individual has the correct (full information) on the true value of being in the R&D sector,V_i^R. But with probability $1 - \lambda$ the individual believes that the chances of innovation are lower than they actually are.[4]

[3] This $\widetilde{w}$ could in principle be negative but the evidence in Stern (2004) suggests that it is positive when looking at multiple job offers for post-doctoral biologists. Scientists "pay" about 15-20% of their entry salaries to be scientists.

[4] Formally, the agent underestimates the extent of the complementarity between innovation and human capital.

The idea is that greater exposure to innovation will increase the chance that an individual believes $\lambda = 1$. If this exposure is greater for some groups than others (such as rich vs. poor), this will be a cause of there being fewer poor (but equally able) inventors.

Define the difference between the perceived value of the R&D sector vs. the non-R&D sector as:

$$\varphi_i = V_i^{PR} - V_i^N = \lambda_i E_\pi(u[T(\rho H_i + \pi H_i)]) + (1-\lambda_i) E_\pi(u[T(\rho H_i + \eta \pi H_i)]) - u[T(\widetilde{w} + \rho H_i)] + \tilde{v}_i$$

with $\tilde{v}_i = v_i^R - v_i^N$.

Whether an individual chooses the R&D sector will depend on the sign of φ_i. If we define $I(\varphi_i)$ an indicator function equal to one if $\varphi_i > 0$ and zero otherwise, then to calculate the number of inventors we simply have to integrate $I(\varphi_i)$ across individuals. To calculate the fraction of a group (low income kids, minorities or women) who become inventors we integrate over the individuals in the relevant group.

The comparative statics of the model are straightforward.

1. Inventor probability increases with the correct information/exposure to the R&D sector

$\frac{\partial \varphi_i}{\partial \lambda_i} = E_\pi(u[T(\rho H_i + \pi H_i)]) - E_\pi(u[T(\rho H_i + \eta)]) > 0$

2. The probability of being an inventor increases with human capital (since $\frac{\partial \varphi_i}{\partial H_i} > 0$ *and* $\lambda \geq 0$)

This is because of the complementarity between human capital and the probability of innovation (i.e. $\pi > 0$). A corollary is that inventor probability also increase with initial ability so long as $\frac{\partial H_i}{\partial a_i} > 0$

3. Inventor probability increases with the relative preference for the R&D sector

If the formulation was ηH_i with $\eta < \pi$ this would be like a "tax" on groups who receive less than their productivity in an occupation. There would still be rational sorting for less informed agents in this world (although less of it) because the very talented would still be prepared to go into the R&D sector as their wage would be sufficiently high to compensate them for the tax.

$\frac{\partial \varphi_i}{\partial \tilde{v}_i} > 0$

4. On the margin, high human capital agents will be more likely to enter the R&D sector as their information improves than low human capital agents

$\frac{\partial^2 \varphi_i}{\partial H_i \partial \lambda_i} > 0$

Corollary A. Groups with higher average values of information, initial ability, human capital and/or a preference for the R&D sector will have a higher fraction of inventors.

Corollary B. A group which is badly informed is less positively selected (in terms of human capital) into the R&D sector

Notice that are two types of welfare gains in the model from increasing λ. First, if there are simply too few inventors from the Social Planner's perspective (e.g. because of knowledge externalities) then a higher λ will help address this. Second, a higher λ should improve the *composition* of inventors as the new inventors will come disproportionately from the high human capital agents. This is the essence of Corollary B which gives a different result from the basic rational sorting model.

To take an example, say that there are two groups, rich and poor. The rich have $\lambda = 1$ and the poor have $\lambda = 0$. Human capital and preferences for working across the sectors are heterogeneous across individuals but identically distributed in the two groups. Since some of the poor will have a strong preference for the R&D sector some will become inventors, but they will be on average of lower human capital compared to inventors from rich families. As we increase λ for the poor, the individuals who choose to enter the R&D sector will be the most talented of the poor. At some point the average human capital of inventors will be the same for the two groups. So an increase in λ not only increases the number of inventors it also increases the average quality, which will have a stronger effect on successful innovation than simply adding another inventor of the same average incumbent ability.

B.1.2 Empirical Implications

We find support for several of the predictions of the model. First, the evidence of the strong correlation between early math test scores and inventor status is consistent with result 2, as is the college-innovation relationship. Second, we find that people more exposed to innovation (even by class of technology) are more likely to become inventors which is consistent with result 1.

A criticism of this result is that the exposure measure may reflect other mechanisms. For example, rather than increasing λ, exposure might work through increasing $\tilde{v}_i$, the preference for an R&D career. In terms of welfare, if the main concern is insufficient numbers of people (of all ability) choosing to be inventors then it does not matter too much whether a policy of exposure works through information or preferences. However, it is unclear that just shifting preferences will have any implications for talent misallocation.

B.2 Extensions

There are many extensions that can be made to the basic model which we now consider.

B.2.1. Endogenous acquisition of Human Capital

In stage 1 we can allow agents to invest in their own human capital. Following Hsieh et al. (2013) we can model this as a "goods tax" (τ_g) on human capital investment that is higher for some groups than others.[5] This is a reduced form way of capturing things like poorer schools in poor neighborhoods. Those who have a low value of λ will rationally choose to invest less in their human capital all else equal. So this will compound the degree of misallocation. Disadvantaged young people perceive a lower return from their talents, invest less and so make their miss-perceptions self-fulfilling. There is evidence that experiments to improve the information of disadvantaged kids can make them more likely to attend college (e.g. Hoxby & Turner (2015); McNally (2013)). In the context of our model this mean they

[5] They also allow for direct discrimination in the labor market whereby some disadvantaged groups obtain a lower wage for their marginal product than others. This is unlikely to be an issue in our context for children from low income families, but it might be for women and minorities. In any case, this turns out to be observationally equivalent to (τ_g).

have a greater chance of becoming an inventor.

B.2.2. Multiple Occupational Sectors

We have simplified our model into having two sectors, but there is no difficulty in allowing multiple sectors. One complication arises between general and occupational specific human capital in such a model, however. In our set-up general human capital gives agents a comparative advantage in the innovation sector. In a multi-occupational Roy model it makes more sense to distinguish between different talents in different sectors. Agents can be born with an initial draw of such talent and will allocate themselves (possibly with endogenous skill acquisition) across these sectors.

B.3 Top Rates of Income Tax: A Simple Calibration

In this sub-section we use the model developed above to consider a quantification of the impact of changing top tax rates on the incentives to become an inventor. We simulate this using our empirical data. Increasing the top tax rate has a benefit that it brings in more revenue to spend on public goods, so when considering such a tax policy change we have to benchmark this in some way to make it revenue neutral. We benchmark against increasing the standard rate of tax so that we consider how raising a dollar through increasing the top tax rate compares with raising a dollar through increasing the standard rate. Our calculations are equivalently to thinking of how cost effective it would be to incentive innovation by reducing the top rate of tax.

B.3.1 Calibration Framework

Step 1. We start with a nonparametric estimate of the pre-tax empirical earnings distribution in the innovation sector. Our proxy for the "permanent income" of an inventor is the average "adjusted gross income" (reported on 1040 Forms) of an inventor between age 40 and 45, minus their spouse's wage income (reported on the spouse's W2 Forms). This measure include the inventor's wage and non-wage income, such as royalties. Our estimate of the empirical earnings distribution is thus simply the percentiles of the observed income distribution. For the rest of the calibration, we work with these 100 cells.

Step 2. Next, we compute expected utility in the innovation sector under various assumptions about the utility function and about the tax regime. Specifically, we consider various CRRA utility functions, $u(c) = \frac{c^{1-\delta}}{1-\delta}$ where c = consumption and the relative risk aversion parameter is δ. We examine δ of 0 (i.e linear utility), 0.5, 1 (i.e. log utility), 1.5 and 2 respectively. We also consider various tax regimes. The status-quo tax regime $\bar{\tau}$ approximates the US tax system, with a tax rate of 28.5% below $439,000 and of 40% above. We then consider a tax regime which keeps the same top tax rate (40%) but increases the standard tax rate by one percentage point to to 29.5% - we will refer to this scenario as the "benchmark policy change" (τ^B) relative to the status quo. We then consider tax regime τ^1 with the same standard tax rate (28.5%) but increasing the top tax by one percentage point to 41%. We will make comparisons between the various tax regimes in terms of "fall in innovation per dollar of revenue raised", so that any increase in top taxes is revenue neutral compared to the alternative policy of simply increasing the standard rate. In other tax policy experiments we consider tax regimes similar to the status-quo tax regime but introducing an additional top tax threshold with a tax rate of 60% beyond this threshold.

We use the estimated empirical earnings distribution (the 100 cells from Step 1) and apply the tax regime to obtain post-tax earnings at each point of the income distribution. We then apply the utility function (assuming that consumption is equal to earnings, i.e. there are no savings) to obtain utility in each of the 100 cells, and finally we average over all cells to obtain expected utility. Likewise, we obtain expected tax revenue under each tax regime.

Step 3. For each tax regime and degree of risk aversion, we compute the "certainty equivalent" (post-tax) wage. The certainty equivalent wage governs the inventor's decision of whether to enter the (risky) R&D sector, as opposed to joining the (safe) non-innovation sector. We assume that there is uncertainty over earnings only in the innovation sector and that each inventor has a "safe" outside option, in the form of a fixed (post-tax) wage in the non-innovation sector. The certainty equivalent is the fixed wage level such that the agent is indifferent between getting this wage for certain (in the non-innovation sector) and drawing from the empirical earnings distribution in the innovation sector. The assumption that tax policy changes do not affect wages in the non-R&D sector is make for analytical simplicity[6] and show in sub-section B.3.3 that this can be relaxed and does not drive our quantitative results (in sub-section B.3.3. we adopt a more parametric approach and specify an earnings distribution for inventors both inside and outside of the R&D sector).

In Step 3, the certainty equivalent is obtained very simply by starting from the expected utility computed in Step 2 and inverting the utility function to recover the certainty equivalent (post-tax) wage.

Step 4. Given the results from step 3, we compute the change in certainty wage equivalent $\frac{dW}{d\tau}$ for each tax regime τ, relative to the status quo tax regime.

Step 5. Finally, we compute the change in the fraction of people becoming inventors in response to a change in taxation. Formally, this is equivalent to the (marginal) deadweight cost of taxation per dollar of tax revenue raised by the government (denoted $\gamma(\tau)$) when switching from the status quo policy $\bar{\tau}$ to the new tax regime τ. We normalize the total labor force to 1 and denote by ϕ the fraction of the labor force choosing to work in the R&D sector. We also compute the expected tax revenue from inventors under each tax regime.

We want to estimate the following quantity:

$\gamma(\tau) = \frac{\frac{d\phi}{d\tau}}{\phi E[R^{\tau} - R^{\bar{\tau}}]} = \epsilon \frac{\frac{dW}{d\tau}}{W^{\bar{\tau}}} \frac{1}{E[R^{\tau} - R^{\bar{\tau}}]}$

where $\frac{d\phi}{d\tau}$ is the change in the fraction of the labor force in the R&D sector as a result of the

[6] In other words, the certainty equivalent responds to tax policy changes only to the extent that they affect expected utility in the innovation sector. This is in line with our general focus on the effect of taxes on inventors and it greatly simplifies the calibration because we do not need to estimate the (counterfactual) earnings distribution the inventors would have obtained had they worked outside of the innovation sector. One way to motivate this assumption is that the changes in tax policy we consider are in a range that is well above the certainty equivalent wage for inventors.

policy change, and $R^{\tau}$and $R^{\bar{\tau}}$ are the amount of revenue raised under the new and status quo tax regimes, respectively. The expectation is taken with respect to the empirical earnings distribution. $\frac{\frac{dW}{d\tau}}{W^{\bar{\tau}}}$ is the percentage change in the certainty equivalent wage (W) between the new tax regime and the status quo, and ϵ is the elasticity of occupational choice with respect to the change in the certainty equivalent. By definition, $\epsilon = \left(\frac{\frac{d\phi}{d\tau}}{\phi}\right) / \left(\frac{\frac{dW}{d\tau}}{W^{\bar{\tau}}}\right)$, i.e. if the certainty equivalent wage decreases by 1% in the new tax regime, the fraction of the labor force going into innovation decreases by ϵ%.

As is standard in public finance, the deadweight cost (innovation impact in our application) of tax crucially depends on a behavioral elasticity, here denoted ϵ. To make the point that $\gamma(\tau)$ is small for increases in top tax rates, we do not need to estimate the value of ϵ. Instead, we express everything *relative to the benchmark policy change*, denoted τ^B (increasing the tax rate below \$439,000 from 28.5% to 29.5%). Under the assumption of a constant elasticity, we obtain:

$$\frac{\gamma(\tau)}{\gamma(\tau^B)} = \frac{\frac{dW}{d\tau}}{\frac{dW}{d\tau^B}} \frac{E[R^{\tau^B} - R^{\bar{\tau}}]}{E[R^{\tau} - R^{\bar{\tau}}]}$$

We know $\frac{\frac{dW}{d\tau}}{\frac{dW}{d\tau^B}}$ from Step 3 and $\frac{E[R^{\tau^B} - R^{\bar{\tau}}]}{E[R^{\tau} - R^{\bar{\tau}}]}$ from Step 2. Hence, we can summarize the impact of a top tax change on the amount of inventors, by examining how the tax change effects the utility (certainty equivalent) of going into the innovation sector.

To summarize, the calibration is based on the following steps:

1. Get the 100 percentiles of the distribution of earnings for inventors between age 40 and 45 from 1040 Forms.

2. Calculate expected utility and expected tax revenue in the innovation sector under the various tax regimes and utility functions.

3. Calculate the certainty equivalent (post-tax) wage under the various tax regimes and utility functions.

4. Compute the change in the certainty equivalent (post-tax) wage relative to the status quo, for each tax regime and utility function.

5. Using the previous estimates, compute the fall in innovation per dollar of tax revenue raised for each tax policy change, relative to the benchmark policy change. Repeat for the various utility functions.

Sub-section B.3.2 below reports these results and discusses the role of risk aversion. Section B.3.3 moves to a more parametric framework to illustrate the role of skewness of the earnings distribution in generating these results.

B.3.2 Magnitude of the Innovation effect and The Role of Risk Aversion

Figure 17 reports the value of the ratio $\frac{\gamma(\tau^1)}{\gamma(\tau^B)}$, defined as above, where τ^1 is the tax policy regime increasing the top tax rate, but keeping the standard rate fixed (i.e. tax rate of 28.5% below \$439,000 and a 41% tax rate above this threshold). Under risk neutrality (linear utility), by definition the two policies have the same deadweight cost because the behavioral response is the same (the inventor values an additional dollar the same at any point of the earnings distribution). The figure shows that the relative efficiency cost steeply declines with the coefficient of relative risk aversion. For standard values of the coefficient of relative risk aversion above 1, the efficiency cost of the increase in top tax rates is one order of magnitude smaller than the efficiency cost of the benchmark policy change. For a coefficient of relative risk aversion of 2, there is essentially no effect on innovation.

Figure A18 illustrates the innovation loss relative to the benchmark for a number of other policy changes, increasing the top tax rate above \$439,000 to 41%, 45%, 50%, 60%, etc. up to 95%. The innovation loss is of course increasing in the tax rate, but the figure shows that the magnitude of the efficiency loss always remains very small (below 12% of the loss in the benchmark policy change) for reasonable values of the coefficient of relative risk aversion, above 1.

Figure A19 takes this point one step further by showing the effect of introducing a new top tax bracket with a 60% top tax rate above a variety of thresholds ($439,000, $800,000, $1.6m, $4m, $10m, $20m, and $30m). The figure shows that "millionaire taxes" have extremely small innovation losses, equal to at most a couple of percents of the innovation loss of the benchmark policy change. The intuition behind this result is that changing the probability of extremely high payoffs does not affect the certainty equivalent wage by much, due to the concavity of the utility function.

All of this results are quantitatively equivalent when winsorizing the inventors' empirical earnings distribution to $100,000. In other words, the results are not driven by the fact that the benchmark policy would have a comparatively large effect because it increases the tax rates on much lower income levels (which are very rare levels of income in the population of inventors).

B.3.3 The Role of Skewness of income returns to being an inventor

To investigate the role of skewness in more detail we adopt a more parametric approach, specifying an earnings distribution for inventors outside of the R&D sectors, in order to show that for very skewed pay-off function, taxes matter less than for less skewed ones in the presence of concave utility. Consider lowering marginal tax rates on high income earners. Assume that the CDF of the returns to innovation is $F(\pi)$ and the CDF of the taste for the two sectors is $G(v)$. Since the baseline wage is common in both sectors except for a shift factor that is common across individuals (above denoted $\widetilde{w}$),we abstract away from this. The benefit of working in the R&D sector is:

$\varphi = F_0(\pi, v) = F(\pi(1-\tau)) - G(v)$

Consider the effect of changing tax on the probability of going into the R&D sector:

$\frac{\partial \phi}{\partial(1-\tau)} = f_0 E[u'(\pi(1-t))\pi]$

$= f_0 \frac{E[u'\pi]}{E\pi} E\pi = f_0 \bar{u}' I$

where f_0 is the PDF of F_0 and $\bar{u}'$ is the profit weighted mean of marginal utility (u'). When $F(\pi(1-\tau))$ gets more skewed $\bar{u}'$ will get small, so the marginal effects of tax on entering the R&D sector will also become small holding innovation revenue (I) fixed.

The key aspect of the result is that for very skewed pay-off functions, taxes will matter less than for less skewed ones in the presence of concave utility. To examine this further we calibrate the model to some empirical distributions. To do this we put some more structure on the model. We assume that innovation returns are Pareto distributed, tastes are log normally distributed and utility takes the CRRA form (as above).We draw a million individuals from these distributions for different levels of the Pareto tail parameter, α between 1.1 and 2. This determines the skewness of the distribution ($\frac{\alpha}{\alpha-1}$), with lower levels denoting a more skewed ("thick tailed") distribution. We then consider the effects of a one percentage point change in the top marginal tax rate from its current level US level (as discussed above). The mean effect of lowering marginal taxation for different levels of skewness of the innovation returns is shown in Figure A20 (where we have used a risk aversion parameter to be δ =1.5).

Lowering taxes always has a positive effect on innovation which is why all the values on the y-axis are above zero. R&D falls by about 1% when skewness is 3 ($\alpha = 1.5$) , for example. However, as skewness increases, the marginal effect of taxation falls. When skewness is 5 ($\alpha = 1.25$) the marginal effect of innovation is about -0.20% and effectively zero for very high levels of skewness. Our analysis of inventor careers showed heavy levels of skewness is consistent with other papers on the distribution of patent values using other methods of valuing patents such as future citations, patent renewal fees, licensing revenues or surveys of inventors.

B.4 Gifted and Talented Programs for the Disadvantaged

A supply side policy would be to increase human capital so that there are more highly skilled people who could become inventors. We have seen that children born to poor parents appear to be at a particular disadvantage in growing up to become inventors in later life and have argued that this is, in large part due to their slower acquisition of human capital during school years rather than "initial ability". Test scores at grade 3 only accounted for 30% of the lower probability of children from poorer families growing up to be inventors, whereas education by 22 years of age could account for virtually all of the income-invention gap.

Card & Giulano (2014) finds evidence that a Gifted and Talented program particularly benefited high ability children from lower income families. Tracking such children and putting them in separate classrooms within public schools raised math and reading by 0.3 of a standard deviation. Looking at the evidence from Figure 3, this implies that such a program would roughly double the future invention rate among the treated group. Since the cost of this program was effectively zero, such policies would seem very desirable on grounds of both equity and growth.[7] By contrast, reducing top marginal tax rates is likely to have some cost in terms of lost revenue to the government.

The pay-off in terms of innovation for such educational programs are long-term. The impact of cutting marginal tax rates is somewhat speedier, although note that our model assumes that this would effect the flow of new graduates into the R&D sector, rather than immediately affecting the stock of inventors (it is different than an R&D tax credit in this respect).

B.5 A Simple Calibration of the Effect of Educational Supply-Side Policies

As discussed in Section VI.C, our estimates can be used to assess the potential gains from supply-side ("extensive margin") policies. Three sets of policy parameters are needed to compute the effect of such policies on innovation. First, we must determine a "reference income group" for which the rate of innovation is policy invariant. Both intuitively and on the basis of the evidence presented in Card & Giulano (2014), supply-side policies will mostly affect disadvantaged students. Accordingly, we consider three policy scenarios with

[7] Another set of supply side policies would be the spreading of high performance school practices as detailed by Fryer (2014) and many others. These seem to have particular benefits for disadvantaged children.

various reference income groups: the policy will reduce a certain fraction of the innovation gap between these (high-income) groups and the rest of the population. Supply-side policies affect only students in families below the 90th percentile of the income distribution in the first scenario we consider, below the 80th percentile in the second scenario, and below the median in the third scenario. Second, we must determine the fraction of the innovation gap between the reference group and the rest of the population that is due to ability. We consider a scenario in which ability accounts for 30% of the gap (in line with our empirical estimates of Table 1) and another scenario in which it accounts for half of the gap (which we view as an upper bound). Finally, we must determine the fraction of the innovation gap - after adjusting for ability - that can be closed by supply side policies. We consider a series of scenarios in which the innovation gap between the reference group and the rest of the population can be closed by between 5% and 50%, respectively.

Figure D.9.: A Calibration of the Potential Gains from Supply-Side Policies (% Increase in Inventor Population)

Reference Group:		**Richest 10%**		**Richest 20%**		**Richest 50%**	
Share of Innovation Gap From Ability:		**30%**	**50%**	**30%**	**50%**	**30%**	**50%**
	5%	7.5%	5.4%	5.2%	3.7%	2.1%	1.5%
	10%	15.1%	10.8%	10.5%	7.5%	4.2%	3.0%
Fraction of the Ability-Adjusted	**20%**	30.2%	21.6%	21.1%	15.1%	8.5%	6.1%
Innovation Gap Closed by Policy:	**30%**	45.4%	32.4%	31.6%	22.6%	12.7%	9.1%
	40%	60.5%	43.2%	42.2%	30.1%	16.9%	12.1%
	50%	75.6%	54.0%	52.7%	37.7%	21.2%	15.1%

Appendix Table B1 reports the percentage increase in the population of inventors under these various scenarios - the gains are in general very large. For each percentile of the income distribution, we compute the implied increase in the percentage of inventors and then average over the relevant range of the income distribution. Alternatively, for each scenario the calculation can be carried out in one step as follows:$\%\Delta Inventors = \frac{s\times(1-a)\times(R^I-R^{\bar{I}})\times(1-P^I)}{R^I\times P^I+R^{\bar{I}}\times(1-P^I)} \times 100$

where s is the share of the (ability-adjusted) innovation gap closed by the policy, a is the share of the initial innovation gap accounted for by ability (which by definition cannot be closed by policy), R^I is the average rate of inventors in reference income group I, $R^{\bar{I}}$ is the average rate of inventors in the rest of the population, and P^I is the share of the reference income group in the total population.

The benchmark scenario discussed in Section VI.C uses the top 10% as the reference group, assumes that ability accounts for 30% of the innovation gap, and that policy can close 20% of the ability-adjusted innovation gap. The increase in the number of inventors induced by the policy is equal to over 30% of the inventor population.

Another useful benchmark to consider is based on the distribution of the innovation-income gaps across states (where we consider state of birth). On average, the rate of inventors among children in the top 20% of the family income distribution is 177% higher than for children in the bottom 80%, but there is a lot of variation across states. The 5th percentile of the distribution of innovation gaps across states is 132% (e.g. New Hampshire) while the 95th percentile is 222% (e.g. Georgia). The 5th percentile of the distribution can be considered to be a "feasible benchmark" that other states could potentially converge to.[8] In other words, the innovation-income gap could be reduced by $\frac{177-132}{177} = 25.4\%$. Using a calculation similar to above[9], this corresponds to a 38.4% increase in the overall inventor population.

[8]We have checked that, when expressed in percentages, the state income-innovation gap is not correlated with either the state population or with the number of patents per resident. These results are available from the authors upon request.

[9]The formula is

$$\frac{0.254 \times (R^I - R^{\bar{I}}) \times (1 - P^I)}{R^I \times P^I + R^{\bar{I}} \times (1 - P^I)} \times 100$$

Regarding the composition effect, the data underlying Figure 2 shows that there is a 44% gap between the 3rd grade test scores (expressed in standard deviations relative to the mean) of inventors from high income families and those of inventors from families below the 80th percentile in the income distribution. In Section VI.A, we have discussed that this difference can result from the fact that high-ability children from low-income families are less likely to enter the R&D sector than their high-income counterparts because they have had less exposure to innovation. One way of calibrating the magnitude of the composition effect is to consider hypothetical policies that would keep constant the total number of inventors from low-income families going into the R&D sector but that would bring into this sector a higher share of high-ability low-income children and a lower share of low-ability low-income, compared to the current equilibrium. The composition can have an effect on the overall rate of innovation (and growth) since higher ability individuals will produce better innovations. The details of the calculation are reported below. We estimate the relevant parameters and find that the composition effect is smaller than the level effect. The reason for this is that the level effect is very large: intuitively, we found that innovation rates are very different across groups (a tenfold difference), while test scores conditional on innovation (innate ability of individuals going into innovation) are much more similar.

The magnitude of the composition effect can be calibrated as follows:

$$\%\Delta Innovation = \frac{\widetilde{s}\times(Q^I - Q^{\bar{I}})}{Q^{\bar{I}}} \cdot \frac{Q^{\bar{I}}\times(1-P^I)}{Q^I\times P^I + Q^{\bar{I}}\times(1-P^I)} \times 100$$

where Q^I is the expected amount of (quality-adjusted) innovations over the course of the career of an inventor in the reference income group I ($Q^{\bar{I}}$ is the same in the rest of the population) and $\widetilde{s}$ is the share of the "quality-adjusted innovation gap" $(Q^I - Q^{\bar{I}})$ between the two income groups that can be closed by policy. Thus, the first term in the formula is the share of the quality-adjusted innovation gap that can be closed by policy and the second term is the share of total quality-adjusted innovations that is accounted for by the

where I is the top 20% and $\bar{I}$ the bottom 80%.

low-income group. Consider an example where the reference group I is children from families in the top 10% of the income distribution. In the data, I represents 34% of quality-adjusted innovation (as measured by citation-weighted patents). We also observe that these inventors have third-grade test scores that are 44% higher than inventors in the rest of the distribution. Under the assumptions (i) that a percent increase in third-grade test score corresponds to a one percent increase in expected quality-adjusted innovations during an inventor's career, and (ii) that supply-side policies can close 20% of the gap in third-grade test scores, then such policies would increase innovation by just under 6% ($0.20 \times 0.44 \times (1 - 0.34) = 5.81\%$).