Essays on College Majors and the Labor Market

by

Shawn M. Martin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in The University of Michigan
2022

Doctoral Committee:

Professor Brian A. Jacob, Chair
Assistant Professor Michael Mueller-Smith
Associate Professor Kevin Stange
Professor Basit Zafar

Shawn M. Martin
shawnmm@umich.edu
ORCID ID 0000-0002-2745-3048

To my parents, Jeanette and William,
and to my partner, Paul

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help and support of my dissertation committee: Brian Jacob, Kevin Stange, Michael Mueller-Smith and Basit Zafar. Brian Jacob has guided me from my first year at Michigan and has provided continuous support. He is patient with me and always willing to talk through details, however small, but somehow still encourages me to see the forest through the trees. Kevin Stange's enthusiasm for research is inspiring and has encouraged me in the times I struggled the most. He is also a wonderful co-author who has provided me opportunities to develop as a researcher. Mike Mueller-Smith inspired me to aim high and take risks; without his support and the examples he sets in his own work, I would have never applied for and succeeded in using the data crucial to the first chapter of my dissertation. Basit Zafar's guidance reminded me to see the bigger picture and think critically about how my research contributes to economics.

Many other faculty at the University of Michigan were important to my development. Charlie Brown provided helpful guidance to me on a range of projects, was generous with his time, and always asked the question you hoped wouldn't be asked but should know the answer to. I also benefited from interactions with John Bound, Susan Dynarski, Sara Heller and Chris Weiland. Numerous individuals have provided key support in navigating other aspects of my graduate training including Hiba Baghdadi, Jasmina Camo-Biogradlija, Kathryn Cardenas, Laura Flak, Julie Heintz, Kyle Kwaiser, Mim Jones, and Nicole Wagner. Outside of the University of Michigan, I have learned a lot from working with the co-authors of the second chapter of my dissertation: Steven W. Hemelt and Brad J. Hershbein. Working with the other members of the larger research team affiliated with the second chapter – including Johnathon G. Conzelmann, Andrew R. Simon and Kevin Stange – has shown me what an effective and enjoyable research partnerships look like.

I am also indebted to several government agencies. I am grateful to the Institution for Education Sciences at the U.S. Department of Education, which funded my research and graduate training through the Causal Inference in Education Policy Research (CIEPR) Predoctoral Fellowship (PR/Award R305B150012#). I am grateful to the U.S. Census Bureau and the Internal Revenue Service, which provided me with access to incredibly rich administrative data. I appreciate the time and service of the Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers, who reviewed all results for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to the released results.[1] Graduate students at the University of Michigan are incredibly fortunate to have on-campus access to a Federal Statistical Research Data Center (RDC), a place that served as my effective office for the last two years. The RDC administrator, Jack Carter, provided crucial support and assistance, and always displayed kindness and patience in working with me. Andrew Foote also provided invaluable guidance on navigating the disclosure avoidance procedures.

I would not have been able to attend the University of Michigan without the support of several other individuals. Roger H. Gordon at the University of California-San Diego, guided me through my first empirical economics research project and motivated me to pursue doctoral studies. During my time at the Office of the Comptroller of the Currency, Phillip Li and Thomas Mayock took me under their wing and shaped my early quantitative research skills.

I would also like to thank my Education Policy Initiative (EPI) classmates and officemates including Elizabeth Burland, Max Gross, Stephanie Owen, Shwetha Raghuraman, Michael Ricks, Andrew Simon, Brittany Vasquez, and Hannah Zlotnick. I am grateful to other classmates who have helped me along the way including Sam Haltenhof, Dylan Moore, Michael Murto, Evelyn Smith and Dyanne Vaught. I am especially grateful to my close friends in Ann Arbor including Jennifer Mayo, Naomi Rawitz, Marco Rojas and Samuel Stern. I am also indebted to my therapist Brianna Colbert who couched me through the last two years of the PhD.

I could not have written this dissertation without the support of my family including my siblings Kristal, Marlena, Nicole and Theo, and my grandmas Barbara and Nelly. My parents Jeanette and William have supported and encouraged me on every step of this journey, and I owe so much of what I have achieved to them. Hannah Kerne has been a wonderful friend, who both inspires and grounds me, since we met ten years ago in rural China. Last but by far not least, I am grateful for the emotional support from and countless laughs shared with my partner Paul Kindsgrab, who made these last six years so much brighter.

[1]The research for Chapter 1 was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2420. Disclosure avoidance approval numbers include CBDRB-FY21-P2420 -R8935 & -R9067 and CBDRB-FY22-P2420 -R9230 & -R9257 & -R9606). Any views expressed or mistakes are my own, and not those of the U.S. Census Bureau. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vii

LIST OF TABLES ix

LIST OF APPENDICES xi

ABSTRACT xii

CHAPTER

I. Job Search and Earnings Growth: General and Specific Majors . . 1

1.1 Introduction 1
1.2 Data 5
1.2.1 Data Sources 5
1.2.2 Analysis Sample 6
1.2.3 Defining College Major Specificity 9
1.3 Empirical Model 11
1.4 Earnings Growth Differs between General & Specific Majors 13
1.5 Conceptual Framework: Job Choice and College Major Specificity . . 16
1.5.1 Setup and Details 16
1.5.2 Implications 18
1.6 Differences in Job Changing by College Major Specificity 19
1.6.1 Alternative Potential Mechanisms 22
1.7 Earnings Growth Accounted for by Job Changing 26
1.8 Conclusion 28

II. College Majors and Skills: Evidence from the Universe of Online Job Ads with Brad Hershbein, Steve Hemelt and Kevin Stange . . 39

2.1 Introduction 39

2.2 Data and Samples 42
2.2.1 Job Ad Data 42
2.2.2 Sample 43
2.2.3 College Majors 44
2.2.4 Categorizing Skills 46
2.2.5 Inferring Desired Skills from Co-Listing with Majors 47
2.2.6 Earnings by Major 48
2.3 Skills Associated with College Majors 49
2.3.1 Measuring Skill Content 49
2.3.2 Skill Specificity of College Majors Based on Location Quotient 51
2.3.3 Measuring Specificity with All Skills 52
2.3.4 Comparison to Prior Work on College Major Specificity . . 53
2.4 Skill Variation Across Areas and Earnings Variability 54
2.4.1 Geographic Variation in Skill Demand 54
2.4.2 Skill Demand and Earnings 55
2.5 Conclusion 58

III. Decomposing Earnings Differences between College Majors 77

3.1 Introduction 77
3.2 Data 79
3.2.1 Baseline Model Variables 81
3.2.2 Full Model Variables 81
3.3 Differences in Mean Earnings between Majors 83
3.3.1 Empirical Approach 83
3.3.2 Results 84
3.4 Decomposing Earnings Differences 85
3.4.1 Empirical Approach 85
3.4.2 Results 89
3.4.3 Detailed Decomposition 91
3.5 Conclusion 93

APPENDICES 104

BIBLIOGRAPHY 191

LIST OF FIGURES

Figure

1.1 Employment Share in Major's 3 Largest Occupations 31
1.2 Earnings Differences between General & Specific Majors 32
1.3 Differences in Employment between General & Specific Majors 33
1.4 Job Changing Rates for General & Specific Majors 34
1.5 Gender Differences in the General-Specific Gap 35
1.6 Earnings Growth and Job Changing by College Major Subject Field 36
1.7 Accounting for Earnings Growth Differences between Majors 38
2.1 Most and Least Frequently Demanded Majors 60
2.2 Comparison between Major Share in Ads vs. BA Completions 61
2.3 Skill Composites: Percentage of Unique Job Postings Containing Skill Composite 62
2.4 Distribution of Average Wage Across Majors and Areas 63
2.5 Distribution of of Skill Concentration Across Majors 64
2.6 Skill Concentration for All Majors 65
2.7 Skill Composite vs. Similarity Index Measure of Concentration 66
2.8 Skill Similarity between Each Pair of Majors 67
2.9 Skill Similarity Index vs. Occupational Measure of Concentration 68
2.10 Variation in Cognitive Skill Demand Across MSAs, Business Majors 69
3.1 Distribution of Residualized Major Fixed Effects ($\hat{\beta}_m$) 96
3.2 Subject Field Categories: Distribution of Low- and High-Paying Majors . . 97
3.3 Percent of Earnings Gap Explained by Covariate Types 99
3.4 Occupation Analysis 101
3.5 Employer Characteristics Analysis 102
A.1 Robustness of Earnings Results to Model Specification 117
A.2 Differences in Labor Supply between General and Specific Majors 118
A.3 Employer Attributes & Cumulative Employers 120
A.4 Gender Differences in Specific-General Major Earnings Gap 121
A.5 Graduate Degree Attainment 122
A.6 Sensitivity of Earnings Growth and Gaps to Graduate Education Controls 123
A.7 Sensitivity of Differences in Job Changing between General and Specific Majors to Graduate Education Controls 124
A.8 Earnings Growth and Gaps in the ACS with and without Graduate Education 125
A.9 Earnings Changes and Job Changes 130

A.10 Coverage of LEHD States 137
A.11 Comparing Earnings across Datasets: ACS-LEHD and NSCG-LEHD . . . 149
A.12 Comparing Job Changing across Datasets: ACS-LEHD and NSCG-LEHD . 150
B.1 Keyword (Old) vs Hand-coded (New) Skill Composites - % of Unique Ads . 157
B.2 Skill Specificity of Majors Using Different Methods to Classify Skills 159
B.3 Correlation between Cell-level Skill Demand and Wages 162
C.1 Percent of Workers Supervising Others by Years Since Graduation 180
C.2 Distribution of Institution Type by Standardized Major Fixed Effect 181

LIST OF TABLES

Table

1.1	Individual-Level Summary Statistics for the Analysis Samples	30
1.2	Graduate Education, Earnings Growth and Job Changing	37
2.1	Occupational Distribution by Sample	70
2.2	Skill Composite Definition and Examples	71
2.3	Share of Ads for Select Majors Indicating Demand for Each Skill Composite	73
2.4	Fraction of Variation in Skill Content Explained by Major and Place	74
2.5	Relationship between Skills and MSA-Major Average Earnings	75
3.1	Variance Decomposition of Work Activities Between & Within Occupations	95
3.2	Gelbach Decomposition of Standardized Major Fixed Effect	98
3.3	ANOVA Analysis and F-tests by Covariate Group	100
3.4	Robustness of the Gelbach Decomposition Results	103
A.1	Person-Year Summary Statistics for the Main Analysis Sample	106
A.2	College Major Specificity Measures	107
A.3	Top Three Most Common Occupations for Select Majors	109
A.4	Characteristics of General and Specific Majors	110
A.5	Total Years in Analysis Sample	111
A.6	Composition of General and Specific Majors, by Gender	111
A.7	Regression Coefficients: Log Annual Earnings	112
A.8	Regression Coefficients: Annual Earnings	113
A.9	Regression Coefficients: Employer and Industry Changing	114
A.10	Regression Coefficients: (Unweighted) Occupation Changing	115
A.11	Regression Coefficients: (Weighted) Occupation Changing	116
A.12	Linear Measure of College Major Specificity: Differences from the Mean	119
A.13	Regression Coefficients: Accounting for Earnings Growth Differences between Majors, an Oaxaca-Blinder Decomposition	126
A.14	Regression Coefficients: Year-on-Year Individual Earnings Gains	131
A.15	Regression Coefficients: Year-on-Year Individual Earnings Gains	132
A.16	College Major Subject Fields	135
A.17	Sample Selection: Individual- and Observation-Level Restrictions	140
A.18	Individual-Level Sample Restrictions & Person Level Summary Statistics	141
B.1	Hand-coded vs Keyword Skill Composites	156
B.2	Explained Variation in Whether a Job Posting Lists at least One College Major	163
B.3	Has-Major F-Test	164

B.4 Categorization of 40 Most Frequently Listed Skills 165
B.5 Complete List of Major Aggregates 166
B.6 Top Skills Associated with Three Majors 167
B.7 Share of Ads for Each Major Indicating Demand for Each Skill Composite 168
B.8 Correlation between Different Measures of Major Skill Specificity 170
B.9 Comparison of Major Rankings by Measure of Specificity 171
B.10 Major Specific Skill Similarity Measures 172
B.11 Replication and Extension of Deming and Kahn (2018) 174
B.12 Raw Cell-level Correlations between Social and Cognitive Skill Content and Wages, Robustness 175
C.1 Descriptive Statistics 177
C.2 Three Most Common Occupations by Primary Work Activity 178
C.3 Three Most Common Primary Work Activities by Occupation 179
C.4 Major Fixed Effects & Standardized Major Fixed Effects 182
C.5 Fraction of Variation in Work Activities Explained by Major and Occupation 184

LIST OF APPENDICES

Appendix

A. Appendix to Chapter 1 105

B. Appendix to Chapter 2 151

C. Appendix to Chapter 3 176

ABSTRACT

This dissertation studies the labor market outcomes of college graduates. The choice of college major is one of the most direct ways for college graduates to acquire skills and earnings differences between college majors rival those between high school and college graduates. In each chapter, I combine rigorous descriptive analysis methods with rich administrative data, job postings data and survey data, to study the determinants of earnings differences between college majors.

The first chapter studies whether earnings growth and job changing varies with the specificity of a college major's skills to particular occupations. College majors vary in the depth and breadth of human capital, and some majors are associated with skills that are very specific to particular work settings. I find that specific majors initially earn 18% more than general majors, but general majors experience steeper earnings growth, and within thirteen years the earnings premium of specific majors shrinks to 6%. To analyze job changing patterns, I create a new linkage of two nationally representative surveys and administrative earnings records. I find that general majors switch occupations, employers and industries at least 20-30% more often that specific majors. A decomposition exercise suggests that these differences in employer and industry change rates account for 30-50% of the explainable portion of the earnings-growth difference between majors.

In the second chapter – joint work with Brad Hershbein (Upjohn Institute), Steve Hemelt (UNC Chapel-Hill) and Kevin Stange (University of Michigan) – we document the skill content of college majors as perceived by employers and expressed in the near universe of U.S. online job ads. Social and organizational skills are general in that they are sought by employers of almost all college majors, whereas other skills are more specialized. In turn, general majors – Business and General Engineering – have skill profiles similar to all majors; Nursing and Education are specialized. These cross-major differences in skill profiles explain considerable wage variation, with little role for within-major differences in skills across areas. We conclude that college majors can generally be conceptualized as bundles of aggregate skills that are fairly portable across areas in ways that occupations are not.

In the third chapter, I investigate the extent to which mean earnings differences between college majors can be accounted for by mean differences in a major's typical job attributes.

Using data from the National Survey of College Graduates, I find that accounting for mean differences between majors in job attributes – including occupation, employer attributes, job tasks and job levels – accounts for (statistically) over half of the earnings differences between majors. I then use the Gelbach (2016) decomposition to quantify the contribution of each type of job attribute. Unsurprisingly, over one-third of the overall explained between-majors earnings differences is due to occupation. An additional one-third is accounted for by mean differences in employer attributes (ownership structure and size) and an additional 20% is due to variation in job-specific tasks.

CHAPTER I

Job Search and Earnings Growth: General and Specific Majors

1.1 Introduction

Differences in average earnings by college major are well-documented and substantial, rivaling the mean earnings gap between high school and college graduates (Altonji et al., 2012, 2016a). Moreover, earnings inequality is dynamic as majors differ in earnings growth rates (Hershbein and Kearney, 2014; Andrews et al., 2019). Yet there is less direct evidence on which characteristics of college majors explain differences in earnings growth, or which mechanisms contribute to these differences. Advancing this understanding is crucial as it has important implications for the post-secondary education investments of students and policymakers (Long et al., 2015; Minaya and Scott-Clayton, 2019).[1]

In this article, I study whether earnings growth varies with the specificity of a college major and the importance of job changing in mediating earnings-growth differences. College majors vary in the depth and breadth of human capital investment: some majors are associated with skills that are very specific to particular work settings whereas others have skills that map less directly into particular jobs. As a result, these differences between *specific majors* (e.g. Accounting and Nursing) and *general majors* (e.g. Communications and Economics) could lead to earnings growth differences through the school-to-work transition. In the short-term, specific majors may be able to more quickly identify jobs that use their skills, and therefore experience initial earnings premiums (Robst, 2007; Lemieux, 2014; Kinsler and Pavan, 2015). General majors, however, may need to spend more time searching for a good job match,

[1] Several studies have found that students' choices of major is responsive to labor market outcomes, although the responses are small (Long et al., 2015; Wiswall and Zafar, 2015; Blom et al., 2021). In addition, federal- and state-level policymakers are increasingly advocating for the use of labor market outcomes in the funding formula that allocate post-secondary funding across institutions and programs (Minaya and Scott-Clayton, 2019).

and experience higher earnings growth in the medium-term (Topel and Ward, 1992; Pavan, 2011). Prior work has emphasized the contrast between general and specific education in labor market outcomes (Hanushek et al., 2017). However, less attention has been paid to this contrast for four-year college graduates and between college majors, and there is even less evidence linking these differences to earnings growth or job changing.[2]

I first show that specific majors have higher annual earnings initially, but the earnings premium erodes over time as general majors experience steeper earnings growth. One to two years post graduation, general majors (majors whose graduates are widely dispersed across occupations) earn 18% less than specific majors. By 13-15 years post graduation the gap is 6%, one-third of the baseline value. Over this period, the average earnings of general majors grow by a total of 93% compared to 72% (.66 vs .54 log points) for specific majors.

Why might differences in earnings growth by college major specificity be related to differences in job changing? I develop a simple conceptual framework in which a worker's human capital upon labor market entry is determined by choice of college major. Each college major is associated with occupation-specific skills, and some majors have a more balanced profile of occupation-specific skills than others. Workers receive a wage offer for a particular job, which in turn, depends on their occupation skills and an idiosyncratic job match. Relative to general majors, specific majors are unlikely to switch occupations as they must draw very high values of the idiosyncratic match component in order to switch to jobs in occupations for which they have little occupation-specific skills.

Measuring individual-level job changes in public-use data is difficult, and I overcome this challenge by performing a novel data linkage of administrative earnings records and two nationally representative surveys. I link post-secondary information from the American Community Survey (ACS) and the National Survey of College Graduates (NSCG), to labor market outcomes in the Longitudinal Employer Household Dynamics (LEHD). I create panel datasets of employment outcomes up to fifteen years post-graduation for four-year college graduates from the 1999-2012 graduating cohorts. An individual's earnings and employment records are linked across multiple states, years, employers, industries, and in some cases, occupations.

Using the linked data, I find that general majors switch occupations, employers and industries much more frequently than specific majors. General majors are 13-18 percentage points more likely to switch broad occupations, a rate that is 35-50% higher than specific majors. General majors also switch employer on average 20% more often and broad industry

[2]Two exceptions include Malamud (2010) and Deming and Noray (2020) who study the impact of the timing of college major choice and technological change, respectively, on post-college labor market dynamics. See also, Leighton and Speer (2020), who focus on measuring human capital specificity among college graduates using earnings premia for a major across occupations.

at least 30% more often.

The average differences between general and specific majors in earnings growth and job changing are robust to considerations of alternative mechanisms. First, the average results don't merely reflect differences between technical (STEM) and non-technical (non-STEM) majors as variation in job changing and earnings growth is present across a wide range of subject fields (e.g. Education, Engineering or Humanities). Second, the average results are not attributable to gender differences in college major choice, as within gender the general-specific major differences persist. Third, although general and specific majors obtain different types of graduate degrees, the aggregate pattern of results exists among individuals prior to graduate education and among those who never enroll.

Finally, using a decomposition analysis, I find that differences in job changing are empirically linked to differences in earnings growth between general and specific majors. Accounting for mean differences between general and specific majors in a variety of covariates – including demographics, economic controls and job changing rates – decreases the earnings-growth difference between general and specific majors by 45-60%, depending on the point of the career. Of this, job changing rates explain (statistically) the largest share, accounting for 30-50% of the explained earnings-growth difference.

This article builds upon previous studies that have documented earnings-growth differences between majors (Hershbein and Kearney, 2014; Andrews et al., 2019), by illustrating which characteristics of college majors are associated with earnings growth variation. In particular, I illustrate the salience of college major specificity. Other researchers have noted the trade-off between general and specific human capital in the context of cross-country differences in the provision of specialized education (e.g. apprenticeships and vocational training) and the timing of academic specialization (Hanushek et al., 2017; Malamud, 2010). One of the only articles to study this trade-off among college graduates is Deming and Noray (2020), who highlight the role technological change and skill obsolescence play in mediating the flatter earnings growth of applied (or technical) majors. This article complements their work by highlighting job search as additional avenue by which the specificity of college major is related to the earnings growth of college graduates.[3] Moreover, the concept of a specific major captured in this article encompasses a wide range of majors, not all of which are technical, including Education, Health and Computer Science & Engineering majors.

I also highlight the salience of job changing as a mechanism for earnings growth among college graduates, which I accomplish by performing a new data linkage. Public-use datasets

[3] Deming and Noray (2020) study why the earnings growth of specific majors is relatively flatter and highlight technological change and skill obsolescence as a key mechanism. In contrast, I focus on why the earnings growth of general majors is relatively steeper and highlight job changing as a key mechanism.

which measure the college major of respondents are primarily cross-sectional (and thus prohibit following individuals over time), or are smaller longitudinal studies (and thus have insufficient sample sizes to precisely measure differences between majors). As a result, past studies on the career paths of college graduates are primarily limited to cross-sectional analyses of the sorting of college majors across occupations. In contrast, I show both differences between majors in the frequency of individual-level job changes, and how these patterns evolve over the early career.

Furthermore, by highlighting job search as a key channel by which mid-career earnings inequality between college majors unfolds, this paper draws on a line of work that studies the job search processes of workers (Neal, 1999), the relationship between job mobility and earnings growth (Topel and Ward, 1992; Engbom, 2022), and the nature of job-to-job moves. Similar to Haltiwanger et al. (2018b), I find that job-to-job moves relocate workers from lower- to higher-paying firms. I expand on their findings by showing that movements across heterogenous firms differ between college graduates, as general majors experience larger increases in the average pay and size of their employers relative to specific majors.[4] Moreover, while past empirical and theoretical research focuses on differences between low- and high-skilled workers in job search (Haltiwanger et al., 2018a), I complement this work by highlighting the salience of a different margin of worker heterogeneity: college major and occupation-specific skills. This margin is crucial given increasing returns to skill (Deming, 2017) and large differences across jobs in the skill- and task-content of work (Acemoglu and Autor, 2011; Deming and Kahn, 2018).

Advancing the understanding of earnings inequality among college graduates is critical to post-secondary education investment decisions. Previous empirical research suggests that students' choice of major is somewhat responsive to labor market outcomes (Long et al., 2015; Wiswall and Zafar, 2015; Blom et al., 2021), and students are increasingly able to search for institution- and program-level labor market outcomes using online tools. At the same time, federal- and state-level policymakers are advocating for the use of labor market outcomes in post-secondary funding formulas (Minaya and Scott-Clayton, 2019). The descriptive results in this article suggest that post-secondary investment decisions should not just consider early-career outcomes; rather, as earnings differences between college majors change throughout the career, the time at which outcomes are measured matters. Finally, as there have been substantive increases in the share of the workforce obtaining bachelor's degrees and shifts in skill demand (Autor et al., 2003; Hershbein and Kahn, 2018; Atalay

[4] Another strand of related literature investigates the role of low- and high-paying firms in generating cross-sectional earnings differences between levels of education (Engbom and Moser, 2017; Cardoso et al., 2018) and among college graduates but between college majors (Ost et al., 2019; Huneeus et al., 2021a). This prior work abstracts from earnings growth and the movement of workers across firms with work experience.

et al., 2020; Deming and Noray, 2020), post-secondary investment decisions will potentially have long-lasting impacts on the aggregate productivity of the economy (Hsieh et al., 2019).

The remainder of the article proceeds as follows. Section 1.2 and 1.3 describe the data and empirical model. Section 1.4 presents the earnings growth results. Section 1.5 describes a conceptual framework linking college major specificity and job changing and Section 1.6 presents empirical evidence on job changing. Section 1.7 links differences in earnings growth to differences in job changing, and Section 1.8 concludes.

1.2 Data

1.2.1 Data Sources

I create a new data linkage between two nationally representative surveys, the American Community Survey (ACS) and the National Survey of College Graduates (NSCG), and administration earnings data from the Longitudinal Employer Household Dynamics (LEHD). All three datasets are accessed via a Federal Statistical Research Data Center (FSRDC) maintained by the U.S. Census Bureau.[5] The size and scope of the data linkage allows the measurement of individual-level longitudinal labor market outcomes for a large and representative sample of the college graduate population, analysis that is infeasible using large cross-sectional datasets or small longitudinal surveys.[6]

The ACS is an annual cross-sectional survey of U.S. households (≈3.5 million households) and contains detailed demographic, education and employment information. I utilize the 2009-2019 ACS waves as college major is only available starting in 2009. The 2010-2019 waves of the NSCG are a rotating panel of individuals initially surveyed in the ACS with a bachelor's degree at the time of survey. It contains detailed post-secondary education and employment data. A single individual can be surveyed up to four times in the NSCG (in addition to their initial ACS response).

The LEHD is administrative data consisting of quarterly earnings by job (employer-employee match) for all employment covered by unemployment insurance (UI). It

[5] Disclosure avoidance approval numbers include CBDRB-FY21-P2420 -R8935 & -R9067 and CBDRB-FY22-P2420 -R9230 & -R9257 & -R9606). For some analysis I also use the public-use versions of ACS extracted from the Integrated Public Use Micro-data Series (IPUMS) and the NSCG extracted from the Scientists and Engineers Statistical Data System (SESTAT) (Ruggles et al., 2020).

[6] This data linkage combines the benefits of the large sample sizes of repeated cross-sections like the ACS with the ability to track individual outcomes using smaller longitudinal surveys like the National Longitudinal Survey of Youth (NLSY) and the Baccalaureate & Beyond (B&B). In addition, the datasets track workers across employment in multiple states (23) which is a much greater scope than that in linked data between a single state's post-secondary records and earnings data. Finally, the micro-data from the Post-Secondary Employment Outcomes (PSEO), which consists of national earnings records for college graduates from several states, is currently unavailable to researchers unaffiliated with the U.S. Census Bureau.

captures roughly 96% of all private-sector jobs and excludes some private sector employers including independent contractors, the unincorporated self-employed, and certain types of nonprofit employers (Abowd et al., 2009). LEHD coverage also includes most employees of state and local government but not employees of the federal government. The LEHD measures all wages and earnings covered by UI including employers' payroll, workers on paid sick leave and paid vacation.

I link the individual-level characteristics of four-year college graduates from the ACS and NSCG surveys in the 1999-2012 graduating cohorts to quarterly earnings data in the LEHD from 23 states and years 2000-2014.[7] The linkage between the three datasets is done using the ACS household and individual identifiers (*cmid* and *pnum*), the Census Bureau's unique identifier (the protected identification key or *PIK*) and a Census Bureau crosswalk between the identifiers. The LEHD earnings data can be linked across years, states and employers and yields between 1-15 annual observations per individual. I subset the LEHD data to only include observations for individuals who are surveyed in the ACS and NSCG. I also separately link an individual's responses across the ACS and NSCG surveys which yields an additional short panel of 2-5 observations per individual.

The above linkage of the ACS, NSCG and LEHD results in three datasets: the ACS-LEHD, the NSCG-LEHD and the ACS-NSCG. The ACS-LEHD dataset is used to measure earnings dynamics, employer and industry changes for the entire sample. The main analysis sample consists of almost 400,000 individuals and over 2.5 million person-year observations. I use the NSCG-LEHD dataset to measure similar outcomes but by graduate attainment. The analysis sample consists of over 11,000 individuals and over 70,000 person-year observations. Finally, I use the ACS-NSCG dataset to measure occupation changes for (30,000 individuals and over 60,000 person-year observations). Appendix A.2 provides additional details on the data construction, linkage processes and sample selection.

1.2.2 Analysis Sample

Individual-Level Restrictions The main analysis samples consist of bachelor's degree recipients from the 1999-2012 graduating cohorts who were surveyed in the 2009-2019 ACS or the 2010-2019 NSCG and labor market outcomes from 2000-2014. Outcomes span up to fifteen years post-graduation depending on the graduation cohort. I restrict attention to the

[7] The full LEHD contains information for all 50 states and D.C but I am only granted access to 23 states including AZ CA CO CT DE KS MD ME ND NJ NM NV OH OK PA SC SD TN UT VA WA WI WY. See Appendix Figure A.10 for a map of the covered states. Estimates from the public-use 2009 ACS suggest the 23 states constitute roughly 50% of national employment. Coverage of particular states in the LEHD varies over time and all 23 states have entered the LEHD system by 2000. I use the 2014 LEHD Snapshot accessed January 2021-March 2022 which includes data through 2014.

1999-2012 graduating cohorts to ensure that for all individuals I can capture outcomes in the first two years post graduation. I drop individuals with a missing or imputed education level and college major, the latter of which occurs for approximately 11% of college graduates in the public-use ACS. Finally, individuals are only included in the analysis sample if they meet an employment restriction in the first year post undergraduate graduation (defined as year 200X for a worker who graduated around May 200Y where Y=X-1).[8] The employment restriction is defined as at least three quarters of non-zero LEHD earnings in the year in the 23 covered states. Finally, I take a 55% random sample of all individuals that meet the above individual-level sample restrictions.[9]

Appendix Table A.18 shows that the distribution of observable characteristics, including gender and curriculum-based major groups (e.g. Humanities and Pure Sciences), remains fairly steady across the various individual-level sample restrictions. One notable difference is the mean graduation year: the average year of bachelor's degree is 2006 in the final ACS-LEHD sample is more recent than that among individuals prior to sample restrictions due to this article's focus on the 1999-2012 cohorts. Table 1.1 also shows that the proportion of individuals that are general and specific majors (which I define in the next section) in the analysis sample is similar to that among a nationally representative sample as measured in the public-use ACS: 45% and 29% respectively. The resulting analysis sample is also nationally representative along other observable dimensions.

Observation-Level Restrictions In the LEHD a worker can have multiple employers in a year and will have one annual observation for each employer. I collapse a worker's observations across employers to yield a single annual observation. For each person-year pair, I measure the total number of quarters with non-zero earnings at any employer. To measure total annual earnings, I sum earnings across all employers in a year. Earnings are inflated to 2014 dollars using the Personal Consumption Expenditures (PCE) deflator from the Bureau of Economic Analysis (BEA). For each person-year observation, a worker's employer is identified by the employer's state-level unemployment insurance account (SEIN) (between the notion of a firm and an establishment). If a worker had multiple employers in a year, the employer is that associated with the worker's main job, the job in the year at which they had the highest earnings. I measure the employer's two- and four-digit industry (NAICS) codes which

[8] I follow Altonji et al. (2016b) and use year and quarter of birth to impute graduation year as the year in which an individual was aged 22 or 23. Estimates from the public-use NSCG indicate that age 22 is the modal age at graduation (over 30%), followed by age 21 (22%) and age 23 (17%).

[9] I use a 55% random sample of individuals that survive the sample restrictions (as opposed to 100%) for Census disclosure review reasons. For the ACS-NSCG sample, I include graduating cohorts up to 2015 to accommodate the oversampling of recent college graduates in the NSCG, and the employment restriction is defined as a non-missing response and occupation in their ACS and first NSCG survey.

correspond to the employer's economic sector and primary business activity, respectively. Employer characteristics, including employer size (number of full-quarter employees) and employer pay (average earnings per full-quarter employee) are measured in the first year of the employer-employee relationship and are fixed until separation so that changes in employer characteristics only occur with employer changes. A worker's occupation is measured at both a detailed level (83 categories) and broad level (20 categories).

Job changes are measured using separate outcomes for employer, industry and occupation changes. Employer and industry changes are measured in the LEHD data and are defined as changes from one year prior as the LEHD data are annual. Occupation changes are measured in the ACS and NSCG datasets, and are measured from two years prior as the NSCG is a biennial survey. For person-year observations that come from the NSCG, I measure the year and type of graduate degree (Master's, Professional and Doctorate) for all graduate degrees received. Finally, as the sampling frame of the NSCG tends to emphasize STEM workers, I used adjusted NSCG survey frame sample weights for some analyses.[10]

College majors are classified using 61 detailed college major categories based on the American Community Survey (ACS) college major codes and created by Hemelt et al. (2021a). I further aggregate a few of their codes to adjust for the NSCG college major codes. Appendix Table A.2 provides the full list of college majors.

For the vast majority of the empirical analysis I focus on annual earnings observations that roughly correspond to full-time, full-year employment. To do so, I omit annual observations that correspond to short and sporadic employment relationships or short-term contract employment arrangements (Barth et al., 2021) by restricting attention to annual observations with at least three quarters of non-zero LEHD earnings in the 23 covered states (i.e. the employment restriction from above). This approach is employed by the Census's Post-Secondary Employment Outcomes (PSEO) and other researchers using unemployment insurance (UI) based earnings records (Minaya and Scott-Clayton, 2019; Foote and Stange, 2019), as the LEHD lacks any measure of work intensity (e.g. weeks and hours worked). For some analysis I flag a worker's annual observations in which they had three plus quarters of non-zero earnings *nationally* using the LEHD U.S. Indicators File. This file provides, on a quarterly basis, the number of states in which a worker has non-zero earnings across all 50 states. The full set of individual- and observation-level sample restrictions for each of the three analysis samples are summarized in Appendix Table A.17.

Appendix Table A.1 contains descriptive statistics at the person-year observation level. In the ACS-LEHD, a worker averages $12,410 in quarterly earnings and $48,850 in annual

[10] Sampling weights are adjusted to account for varying sample sizes across surveys while maintaining the relative weights within each survey in a similar fashion to Altonji and Zhong (2021).

earnings. Roughly 4% of observations correspond to years in which a worker had non-zero earnings in multiple of the 23 covered states. On average, 91% of total annual earnings are earned at the worker's main job employer and over 75% of person-year observations have annual earnings from their main job employer in excess of 90%. This suggests that focusing on annual observations in which a worker has three-plus quarters of non-zero earnings successfully isolates employment relationships that are closer to full-time full-year arrangements. A worker is observed with an average of 1.54 employers with non-zero earnings in a year, but only .970 employers in a year at which they received earnings in excess of the prevailing minimum wage x 35 hours x 50 weeks. A majority of observations correspond to years in which a worker was observed working for just one employer (64%). The mean duration between an individual's annual observations is 1.109 years and over 95% of observation pairs are within one year of each other, indicating that almost all change variables correspond to changes over one-year intervals. Workers change main job employer in 26% of observations. Of these changes roughly 20% and 16% are also a change of four- or two-digit industry, respectively.

1.2.3 Defining College Major Specificity

This article contrasts the labor market outcomes of general and specific majors. Specific majors are those that have skills that are tailored to particular work settings (or occupations) whereas general majors have skills that are not. Specific majors should thus be clustered in the occupations for which they have high levels of skills, but general majors should be more dispersed.

I distinguish between specific and general majors using the employment distribution across occupations of each major's recent graduates. I collapse data from the 2009-2019 public-use ACS to weighted employment counts in each major-occupation cell. Workers are restricted to non-enrolled college graduates who are employed one to five years post undergraduate degree and that work at least part-time part-year.[11] For each major, I measure the percent of bachelor's degree recepients employed in each occupation. I then sum the employment shares across the three largest occupations for the major.

Majors with a high percent of recent graduates employed in the major's three largest occupations are considered specific because graduates are clustered in a small number of occupations (Altonji et al., 2012; Blom et al., 2021). General majors have a lower percent as graduates are more widely dispersed across occupations. In the empirical analysis, majors

[11]The part-time part-year restriction is 27-39 weeks and 20 hours/week. I only include graduates from the 1999-2012 graduation cohorts. These sample restrictions are chosen to mimic the sample of individuals in the ACS-LEHD dataset that I will use for the earnings and job change analysis. The measures and ranking of measures are robust to only including full-time full-year workers (40-47 weeks and 30 hours/week). See Appendix Section A.2.1 for more details.

are divided into 3 equally sized groups based on the measure: general majors are the 20 majors with the lowest percent employed in the major's top 3 occupations, specific majors are those with the 20 highest values and all other majors are considered neither general nor specific. The ranking of majors using this measure is highly correlated to the ranking of majors using either the percentage employed in the major's top 5 occupation or a Herfindahl index (corr>.97).[12]

Figure 1.1 displays the occupational dispersion measure for select majors (see Appendix Table A.2 for all majors). The most specific major is Nursing with over 90% of recent graduates employed in the major's most common occupation of Registered Nurses. Another example is Accounting, with 69% of recent graduates employed in the largest occupation of Accountants and Auditors. The most general major is Other Social Sciences with around 10% of graduates working in the three largest occupations for the major. Majors that are neither general nor specific include Psychology, Statistics and Marketing.

An inspection of the largest three occupations for each major confirms that the occupation dispersion measure yields a reasonable classification of which majors provide general or specific skills. In particular, Table A.3 shows that the top three occupations for the vast majority of the 61 major align with ex ante expectations.[13]

Characteristics of General & Specific Majors Appendix Table A.4 provides additional characteristics about general and specific majors. Both general and specific majors are comprised of majors from a variety of curriculum-based categories of majors. Half of specific majors are Computer Science & Engineering majors, but specific majors also include majors from the fields of Health and Education. Many general majors come from the Social Sciences but also come from all fields except Education and Computer Science & Engineering. Roughly half of the 20 majors that are neither general nor specific are Pure Sciences (e.g. Math, Biological, Life and Physical Sciences). Among the 20 most popular bachelor's degrees in terms of recent degrees granted, 12 are general, 4 are specific and 4 are neither general nor specific.

[12]The occupation dispersion major is constructed with a fairly detailed number of occupations. If I use a broader categorization based in the three-digit Standard Occupational Classification (SOC) codes, the ranking of majors is fairly similar. More specifically, I regress the percentage employed in the major's top 3 occupation measured using the detailed occupations codes on the measure using broader SOC occupations. The r-squared from a regression of the measures is .82 unweighted and .91 when majors are weighted by employment. When majors are ranked according to the measures, the r-squared is .69 unweighted and .75 weighted.

[13]One exception is Visual & Performing Arts, in which the top 3 occupations likely reflect dispersion due to a lack of appropriate employment opportunities. As the occupational dispersion major measure reflects both the labor supply decisions of college graduates and the labor demand decisions of employers, the wide dispersion of Visual & Performing Arts graduates could reflect a lack of labor demand for their skills.

Calculations using the public-use NSCG suggest that there are some slight differences in the types of institutions general and specific majors attend. General majors are somewhat more likely than specific majors to attend Research Universities (RI) and Liberal Arts Colleges, but slightly less likely to attend public institutions. General majors come from slightly more educated families, as the percent of general majors with at least one parent with a college degree is larger than specific majors (58% compared to 52%). The proportion of females is slightly higher among specific majors relative to general majors (63% versus 58%) and the composition of curriculum-based categories of majors differs slightly by gender: the *typical* specific major for females is an Education or Health major whereas for males it is a Computer Science & Engineering major (see Appendix Table A.6).

1.3 Empirical Model

To measure differences in the early career labor market outcomes between general and specific majors, I estimate a regression that permits labor market outcomes to differ by major flexibly over the early career:

$$\begin{aligned} Y_{it} = \beta_0 &+ \sum_g \beta_g g_{it} + \sum_m \phi_m(\text{major group}_{im}) + \sum_m \sum_g \beta_{m,g}(g_{it} \cdot \text{major group}_{im}) \\ &+ \delta X_{it} + \psi Z_i + \theta_t + \gamma_{it} + \epsilon_{it} \end{aligned} \tag{1.3.1}$$

In Equation (1.3.1) Y_{it} is a labor market outcome measured in year t for individual i. The analysis period spans 1 to 15 years post graduation and the variable g_{it} indexes years since (undergraduate) graduation bin g. The variables *major group*$_{im}$ are indicators for college major specificity group equal to one if a worker's undergraduate college major is in group m. The three possible groups are general, specific, and not general or specific majors, and general is the omitted category. The terms $g_{it} \cdot majorgroup_{im}$ are interactions of the major group indicators and the years since graduation bins g. Z_i is a vector of individual characteristics including demographic characteristics (female, black, hispanic), five-year graduation cohort fixed effects and region of birth fixed effects. X_{it} is a vector of time-varying controls including an indicator for LEHD earnings in multiple of the 23 covered states. To control for any initial and medium-term adverse effects of graduating into a recession (Oreopoulos et al., 2012; Altonji et al., 2016b), X_{it} also includes the demeaned unemployment rate in the worker's region of birth in the year of undergraduate graduation and interactions of the demeaned unemployment rate with a quadratic in years since graduation.[14] As a result, the point

[14] This specification assumes that any impacts of graduating into a recession are symmetric across college majors. A interesting point of future research is whether or not general or specific majors are differentially

estimates reflect outcomes for workers who graduated from college in a year with a regional unemployment rate equal to the sample average. Finally, X_{it} includes an indicator flagging wether the worker had three rather than four quarters non-zero LEHD earnings in the 23 covered states. The variables γ_{it} and θ_t are state of employment and year fixed effects.

Differences between specific majors and general majors in the omitted (initial) years since graduation bin g are given by ϕ_m. The coefficient β_g measures the change in the outcome for general majors from the initial period to year since graduation bin g, and $\beta_{m,g}$ captures the additional change for specific majors relative to general majors. The gap in the outcome between general and specific majors in period g is $\phi_m + \beta_{m,g}$. The level of the outcome in period g is equal to $\beta_0 + \beta_g$ for general majors and $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$ for specific majors.[15] Changes in the outcome from years since graduation bins g to $g + 1$ is equal to $\beta_{g+1} - \beta_g$ for general majors and $(\beta_{m,g+1} - \beta_{m,g}) + (\beta_{g+1} - \beta_g)$ for specific majors. In most specifications, I index years since graduation using two-year bins so that g_{it} is equal to one if year t is either year since graduation in the bin g.

Point estimates from Equation (1.3.1) do not measure the causal impact of the choice of general or specific major on an individual's labor market. For the comparison of mean outcomes between majors to measure the causal impact of one major relative to another, the key assumption is that the distribution of ability and preferences is the same among individuals in both majors (Altonji and Zhong, 2021; Lovenheim and Smith, 2022). This amounts to a very strong selection on observables assumption: the set of control variables described above are sufficient to account for any average unobservable differences between general and specific majors that are related to college major choice and post-college earnings and job choice. Researchers employing quasi-experimental variation in student assignment to different majors (Hastings et al., 2013; Kirkeboen et al., 2016) have documented a causal link between college major and labor market outcomes. As a result, while I am overestimating the differences in labor market outcomes between general and specific majors, the differences are likely not completely driven by selection.

impacted by labor market conditions. I use unemployment rate in the region of birth because for the vast majority of individuals, I cannot observe where they graduated college. The state at which a worker is observed in the first year post undergraduate degree could differ from the state of college graduation and is potentially endogenous to the unemployment rate at graduation (Wozniak, 2010). Finally, the use of region as opposed to state unemployment rates allows for the possibility that college graduates search in regional rather than state-specific labor markets post graduation.

[15] In practice, I replace the constant β_0 with the sample mean of the outcome for general majors in the initial period.

1.4 Earnings Growth Differs between General & Specific Majors

I measure earnings differences between general and specific majors using Equation (1.3.1) with (log) annual earnings as the outcome, the ACS-LEHD analysis sample and years since graduation indexed by two-year bins. Figure 1.2 presents the regression-adjusted log earnings profile, the earnings gap between general and specific majors and period-specific earnings growth. See Table A.7 for all corresponding regression estimates and standard errors.

Specific majors initially earn more than general majors, but as the earnings of general majors grow more, the initial earnings premium of specific majors decreases substantially. In the first and second year post graduation, specific majors earn roughly 18% ($p<.01$) more than general majors. While general and specific majors display similar earnings growth between 1-2 and 3-4 years post graduation, in most subsequent periods the earnings growth of general majors outpaces specific majors.[16] From 3-4 to 5-6 years post graduation, the earnings of general majors grow by 14% compared to 11% for specific majors, and ranges from 1.7% to 3.5% ($p<.01$) thereafter. By the end of the analysis window, average earnings among general majors grow by 66% compared to 54% for specific majors. As a result of the sizable differences in earnings growth, the initial earnings premium of specific majors falls to 6% ($p<.01$), which is one third of the initial premium. The earnings premium also decreases in dollar terms by roughly 75%, from $7500 to $2200 over the same period (see Appendix Table A.8).

The main specification indexes years since graduation using two-year bins, but the general pattern of results is robust to alternative parameterizations of years since graduation. When two-year bins are used, earnings growth from period $g-1$ (spanning years a and b) to bin g (spanning years c and d) is measured as average earnings in years c and d less average earnings in years a and b. As a result, results from the specification with two-year bins don't approximate, and are larger than, *annual* changes in earnings. However, Appendix Figure A.1 shows that in a model with either a quadratic function of years since graduation or one-year bins, earnings growth differences are smaller but the earnings premium of specific majors follows the same general path. In both specifications, the earnings growth of general majors tends to be 1 percentage point higher than that of specific majors, compared to a gap of 2-3 percentage points in the main specification. One exception is in the early career and the one-year bin model: in the first two years post graduation the earnings growth of specific majors initially exceeds that of general majors.[17]

[16]Whenever I discuss results from specifications with log earnings as the outcome I discuss coefficients in percentage terms without transforming the log coefficients: i.e. a coefficient of .XYZ is discussed as XY.Z%

[17]Figure A.1 also presents estimates of earnings changes from a regression model with two-year bins and *individual-level* changes in earnings as the outcome. If bin g spans years c and d, then earnings growth

The above contrasts the earnings growth of general and specific majors and omits any discussion of the twenty majors that are neither general nor specific. The outcomes for these majors tend to fall between that of general and specific majors, but are much closer in magnitude to general majors than to specific majors. This pattern of results is perhaps not surprising, as the largest amount of variation in the occupational dispersion measure used to categorize majors occurs between general and specific majors. The results for these majors are provided throughout Appendix A.1.

I also compare majors using a linear measure of occupational dispersion (percent of a major's recent graduates in the major's three largest occupations) rather than the three college major specificity groups. More specifically, I estimate Equation (1.3.1) with the demeaned measure of occupation dispersion interacted with years since graduation bins. In each bin, the coefficient on the measure of occupation dispersion divided by ten equals the differential between a major with the average value of the measure and a major with a value that is 10 percentage points above the mean (i.e. .317 compared to .417).[18] Results are presented in Panel A of Appendix Table A.12. Initially, the earnings premium of a major with an occupation dispersion that is 10 percentage points above the mean (i.e. a major that is more specific than the average major) is 3.6% more ($p<.01$) and decrease to .059% ($p<.01$) by 15 years post graduation.

Labor Supply To what extent are differences in earnings growth a function of labor supply differences between majors? In the main analysis I condition annual observations on having non-zero earnings in three plus quarters in the year, and variation in labor supply can occur along two margins. First, individuals can have either three or four quarters of non-zero earnings. Figure 1.3c shows that the probability of this doesn't differ substantively between general and specific majors over the early career.[19] The primary log earnings regression specification adjusts for mean earnings differences between three- and four-quarter earners with an indicator, but Figure 1.3d shows that the evolution of the log earnings gap over time (and thus earnings growth) does not materially change when the control is excluded. Second, conditional on having non-zero earnings in three plus quarters in the year, I am unable to

in period g is the (weighted) average of year-on-year earnings gains from b to c, and from c to d. As this specification is measuring averages of *annual* earnings changes, the point estimates are more similar to the one-year bin and quadratic model.

[18] Across majors the average percent of recent graduates found in the top 3 occupations is .317. The average for general majors is .15, is .54 for specific majors and .26 for majors that are neither general nor specific. The cutoff point for a general major is .19 and for a specific major is .34.

[19] In the first six years post graduation, specific majors are about 1-2 percentage points more likely than general majors to have 4 rather than 3 quarters of non-zero earnings. Although these estimates are statistically different from zero, they are relatively small differentials compared to the baseline of 84% of general majors 1-2 years post graduation that work 4 rather than 3 quarters.

account for differences in weeks or hours worked in the LEHD analysis. I instead use the 2009-2019 public-use ACS to measure differences between majors in the likelihood of working full-time full-year and earnings growth conditional on full-time full-year employment (a sample that should not be extensively affected by variation in hours and weeks worked). Initially, specific majors are more likely to work full-time full-year, but general majors experience larger year-on-year increases in labor supply (see Appendix Figure A.2a and A.2b). As a result, among all non-zero earners, the log earnings premium of specific majors is substantially larger initially. However even among full-time full-year workers, general majors experience higher earnings growth and the earnings premium of specific majors decreases (see Figure 1.3d).[20]

Employment & LEHD Coverage The analysis sample consists of *individuals* that have at least three quarters of non-zero earnings in the 23 LEHD states in the first year post college graduation, and an individual's annual *observations* that meet the same condition in a given year. While attrition from the analysis sample due to movements outside of the 23 states is unavoidable, the extent of attrition is likely less than in other administrative datasets based on a single state's post-secondary data matched to in-state earnings records. The decrease of the specific majors' earnings premium with years since graduation with suffer from downwards bias if majors differ in LEHD over time and specific majors (general) that leave the sample are positively (negatively) selected on earnings (Foote and Stange, 2019).

Differences between general and specific majors in the pattern of analysis sample coverage are limited to the initial periods and don't lead to substantial differences in cumulative employment. To illustrate this, figure 1.3a plots the probability of being in the analysis sample separately by major and year since graduation bin. Among workers who are 3-4 years post graduation, roughly 78% of general and 85% of specific majors have observations in the analysis sample, respectively. In subsequent years, however, the gap remains roughly proportionate at around 6.5 percentage points or (9% higher) for specific majors. In addition, Appendix Table A.5 shows that general and specific only slightly differ in the total years of coverage in the analysis sample.

While I have limited ability to measure what individuals earn outside of the analysis sample, there is suggestive evidence that the differences between general and specific majors may not be large. Figure 1.3b plots for each year since graduation, the proportions of individuals that aren't in the analysis sample but do have three-plus quarters of earnings

[20] Not all individuals that are missing an annual observation have zero earnings in the year. A person-year observation would be excluded from the analysis sample if the worker 1) works in a non-covered employment type (e.g. unincorporated self-employed or federal employment), 2) had three-plus quarters of non-zero LEHD earnings in another state, or 3) the worker has less than two-quarters of non-zero earnings across all LEHD states (this includes zero earners).

in the LEHD *nationally*. Throughout the entirety of the analysis window the proportion is equivalent for general and specific majors.[21]

1.5 Conceptual Framework: Job Choice and College Major Specificity

In this section I describe a simple conceptual framework that illustrates how skill differences between general and specific majors can impact job search and earnings growth. In the model, a worker's human capital upon labor market entry is determined by choice of college major. College major choice is not explicitly modeled, and skill differences can reflect both selection and treatment. Each college major is associated with occupation-specific skills, and some majors have a more balanced profile of occupation-specific skills than others. Workers receive a wage offer for a particular job (employer $\times$ industry $\times$ occupation) equal to the worker's job-specific marginal product. The marginal product is the sum of the worker's occupation skills and an idiosyncratic job match. Job offers arrive at an exogenous rate and job matches are observed immediately, assumptions that imply workers must search for better job matches.

1.5.1 Setup and Details

Each individual enters the labor market with human capital differentiated by their undergraduate college major. In particular, there are O occupations and all individuals i with undergraduate college major m have human capital vector $H_m = (H_{m,1}, H_{m,2}, .., H_{m,O})$. The vector $H_{m,O}$ doesn't vary across individuals within major, but does vary between majors. Some majors, which I refer to as *specific majors*, have training that is more tailored to particular occupations so that $H_{m,o}$ is high for a few occupations but is low for most. Other majors are *general* and $H_{m,o}$ is similar for most occupations.[22]

The model assumes that college major defines a worker's occupation-specific skills at labor market entry, but doesn't explicitly consider the origin of the skill differences. If college

[21] I am able to calculate this using the LEHD U.S. indicators file which provides, for each quarter in a year, the number of states nationally a worker has non-zero earnings. I simply count the number of individuals that have non-zero earnings in at least one state in three quarters of the year. The final proportion is calculated as: P(3+ quarters earn>0 nationally)-P(3+ quarters earn>0 in 23 states) divided by 1-P(3+ quarters earn>0 in 23 states).

[22] Alternatively, occupation-specific skills could be thought of as a vector of capacities in two tasks α and β. Each major has skills α_m and β_m and specific majors either have predominantly α or predominantly β skills whereas general majors have roughly equal amounts of α and β. Each occupation demands α and β in different proportions given by $T_{\alpha,o}$ and $T_{\beta,o}$, respectively. A worker with major m has occupation o skills equal to $H_o = T_{\alpha,o}\alpha_m + T_{\beta,o}\beta_m$. The values of H_o vary more across occupations for specific majors than for general majors.

major choice is exogenous, then skill differences are developed by the course structure and training that is particular to a college major. If college major choice is endogenous, then skill differences also reflect selection into major based on an individual's preferences for and (un)certainty over their desired career path. For example, prospective college students that are certain about their career paths will have strong preferences for particular occupations and pick specific majors, whereas uncertain students may select general majors. As the latter group would likely change jobs more frequently irrespective of the specificity of their college major, endogenous college major choice overstates the differences in job changing and earnings growth that would result if college major choice was exogenous.[23] Past empirical research suggests that both origins of skill differences between majors are salient (Arcidiacono, 2004; Hastings et al., 2013; Kirkeboen et al., 2016).

In each period a worker receives job offer(s) which arrive at some fixed exogenous rate λ which introduces search frictions. As a result, workers must search for better matches because they do not receive all possible job matches at any given time nor do they immediately locate the optimal job upon labor market entry. A job offer is for a particular combination of employer, industry and occupation and is a wage equal to a worker's job-specific marginal productivity.[24] A worker's job-specific marginal product is the sum of the worker's occupation human capital and an idiosyncratic match factor. When a worker and employer meet, college major is observed and a match factor specific to the job and individual is drawn. Job matches are observed immediately and there is no learning about match quality (Jovanovic, 1979). The job match $\theta_{i,j}$ is draw from the distribution $F(\theta)$ which is i.i.d across workers, implying that $\theta_{i,j}$ is not correlated across occupations. Individual i with college major m has job-specific productivity for job j that is in occupation o equal to $X_{i,m,o,j} = H_{m,o} + \theta_{i,j}$.

This specification of job-specific productivity allows two workers with the same productivity for a given occupation (i.e. same college major) to have different individual-level productivities in the *same job* (i.e $X_{i,m,o,j} \neq X_{k,m.o,j}$ for individuals $i \neq k$). Similarly, a single individual can be more productive working at a job j' in a large firm than in a job j'' at a small firm even in the *same occupation and industry* (i.e. $X_{i,m,o,j'} \neq X_{i,m,o,j''}$ for two jobs j' and j''). For example, Computer Programming jobs require more social or client engagement skills in large firms and client facing industries, and not all Computer Science majors have the same level of social skills.

[23] The model also considers the case of a single major. Assuming that students have some fixed amount of time to study, double majors impact the level of occupation skills a worker can achieve from each of the two majors. A measure of one-dimensional ability a that captures the speed of learning, or the quality of an individual's post-secondary institution could also be incorporate: $H_m = (aH_{m,1}, aH_{m,2}, .., aH_{m,O})$ Selection into college major based on pre-college aptitudes would imply that a varies across occupations.

[24] As there is no labor supply in the model, earnings and wages are equivalent and I use the terms interchangeably.

The above definition of the job-specific productivity also captures the notion that occupation doesn't fully characterize job tasks. There are large differences across occupations in the tasks performed by workers (Acemoglu and Autor, 2011) but employer skill demand (as measured on job postings) also varies within narrowly defined occupations and industries (Deming and Kahn, 2018). Productivity differences across firms related to firm-specific production processes or technology adoption can also impact a worker's productivity (Syverson, 2011).

Each period a worker picks the job that provides the highest earnings and only switches jobs if they receive a better wage offer. A worker compares up to three wages: their wage in their current job j in occupation O, the wage for another job j' in the same occupation O, and the wage for a third job j'' in a different occupation O'. As a result, a worker can either (i) not change jobs, (ii) change jobs within occupation, or (iii) change jobs between occupations. The period-specific maximization problem is summarized by:

$$\max\{ \underbrace{H_{m,O} + \theta_{i,j}}_{\text{current job } j \text{ in occupation } O}, \underbrace{H_{m,O} + \theta_{i,j'}}_{\text{new job } j' \text{ in occupation } O}, \underbrace{H_{m,O'} + \theta_{i,j''}}_{\text{new job } j'' \text{ in occupation } O'} \} \quad (1.5.1)$$

A worker's choice of jobs within occupation only depends on the relative magnitude of their job matches $\theta_{i,j}$, but the choice of jobs between occupation also depends on the relative magnitude of their occupation skills $H_{m,O}$. Finally, as workers only switch jobs if they receive a better wage offer, earnings growth comes from job switching.

1.5.2 Implications

In this framework outlined above, differences in labor market outcomes between general and specific majors will be driven by differences in the probability of *between* occupation changes.

First, general and specific majors are equally likely to change jobs *within* occupations, a choice that only depends on the relative magnitude of job matches. Specifically, a worker changes from j to j' within occupation O if $(\theta_{i,j'} - \theta_{i,j}) > \epsilon$, which is equally likely for general and specific majors (conditional on $\theta_{i,j}$).

Second, the probability of *between* occupation changes differs between general and specific majors, and depends on the current occupation. To illustrate this, suppose that specific majors are highly skilled in occupation O so that their occupation skills satisfy $H_{m,O} >> H_{m,O'}$ for all $O' \neq O$, whereas general majors are equally skilled in most occupations and $H_{m,O} \approx H_{m,O'}$ for all $O' \neq O$. For general majors, changing jobs across occupation, either from O' to O or from O to O', only necessitates a slightly higher draw of a job match. For specific majors, however, the probability of moving from O' to O is not equal to moving from O to O'. The probability of a O' to O move is high, as only a very small job match is necessary to make a job in an occupation for which they have skills (O) more attractive than a job in an occupation for which they don't have skills (O'). In contrast, a O to O' move is unlikely as it requires a job match draw in occupation O' that is sufficiently high to offset forfeiting a high level of occupation O skills.[25] Between occupation changes should be unlikely for specific majors because, intuitively, the probability of a specific major starting their employment in occupation O should be much higher than in O'. In totality, the balance of job switches between general and specific majors will depend on the initial distribution of occupations and job matches.

In the model, job changing differences are driven by occupation changing, but in the data, I can only measure occupation changes for a subset of individuals (the ACS-NSCG dataset). As employer and industry changes observed in the ACS-LEHD could either be job changes that occur within or between occupations, I will also explore between-major differences in employer and industry changes. Finally, the above framework highlights the relationship between college major specificity, job switching and earnings growth. There are likely other characteristics of workers and their labor market experiences that mediate between-majors earnings inequality.

1.6 Differences in Job Changing by College Major Specificity

The conceptual framework described in Section 1.5 suggests that general majors should change jobs more often than specific majors. To test this prediction, I estimate Equation (1.3.1) with indicators for individual-level changes of occupation, employer and industry as the outcome and years since graduation indexed by two-year bins. Industry and employer changes are estimated using the ACS-LEHD dataset and occupation changes are estimated using the ACS-NCSG.

First, general majors are considerably more likely to change occupations than specific majors. Figure 1.4a plots the percent of specific and general majors that switch broad occupation, of which there are about 20 categories, from two years prior, and Figure 1.4d plots the differential between majors. Three to four years after graduation, 57% of general majors changed occupations compared to 38% of specific majors, a differential of 18 percentage points (p< .01) that represents a 50% higher rate. For all workers, the propensity to change occupations decreases monotonically over the early career but the gaps between majors

[25] For example, for a Nursing major to work as a Financial Analyst, a job for which a Nursing major has a presumably low level of occupation skills, the individual must draw an offer for a job as a Financial Analyst with a very high job match in order to be willing to forfeit their Nursing occupation skills.

remain sizable even 13-15 years post graduation: 41% of general majors changed occupations compared to 28% of specific majors, a differential of 13 percentage points (p< .01) that represents a 46% higher rate.

The sizable differences in occupation changes are also present, although are less dramatic, for a narrow classification of occupation (of which there are around 80 categories). The gaps between general and specific majors are statistically significant and range from 9-20 percentage points (p<.01) which represents rates that are 20-40% higher depending on the period. For the unweighted sample, the differential between majors in broad occupation changes is even wider that in the unweighted sample: in some periods the rates of general majors are almost double that of specific majors. See Appendix Tables A.11 and A.10 for all occupation change regression results.

Second, results in Figure 1.4b show that although the gaps between majors in employer changers are relatively smaller than occupation changes, they are still sizable. For example, 33% of general majors changed employers 3-4 years after graduation compared to 27% of specific majors, a differential of 5.8% percentage points (p<.01). As with occupation changes, the propensity to change employers gradually decreases as workers progress through their careers, but the gaps between general and specific majors remain statistically significant and substantive: in almost all periods the rates of general majors are 16-24% higher than specific majors (see Panel A of Appendix Table A.9). Finally, Appendix Figure A.3c shows that differences in employer changes are not just driven by a subset of frequently changing graduates with general majors, rather the distribution of cumulative employers (and thus cumulative employer changes) for general majors is a rightward shift of that for specific majors.

Third, I find suggestive evidence that, on average, employer changes move individuals towards higher paying and larger firms, and more interestingly, that the paths of general and specific majors through heterogeneous firms differ. Appendix Figure A.3 plots the coefficients from separate estimations of Equation (1.3.1) with the outcomes of log employer size and log employer pay. Employer characteristics are measured in the first year in which the worker joined the firm, and remain fixed for the duration of the employer-employee relationship; the characteristics of a worker's employer only change when the worker changes employers.[26] Specific majors initially worker at high-paying firms, but, as general majors switch employers more frequently, they switch to higher-paying firms and the gap in log employer pay shrinks almost entirely from 10% initially. Initial differences between majors in log employer size

[26]The measured changes in employer attributes are likely not mechanical, as employer characteristics are constructed using the universe of workers covered by unemployment insurance records, not just the individuals included in the analysis sample. In addition, while most firms are small, most employment is at large firms, firms for which the addition of a single worker likely has a negligible impact on employer characteristics.

are more dramatic, and although the more frequent job changes of general majors leads to an increase in average firm size, the gap between majors remains sizable even 15 years post graduation. Finally, these results are consistent with variation in the sorting of workers across employer attributes by education levels and college majors documented in prior work (Engbom and Moser, 2017; Cardoso et al., 2018; Ost et al., 2019; Huneeus et al., 2021a).

Fourth, general majors are also more likely to change industries than specific majors (see Figure 1.4c). The percentage point difference between general and specific majors is on par with employer changes, but it represents a larger percent differential. More specifically, among workers in the initial phases of their career 19% of specific majors changed industries, but general majors are 6 percentage points ($p<.01$), or 31%, more likely to change. In almost all subsequent periods general majors change broad industries 27-40% more often than specific majors. Note that the smaller magnitude of industry changes relative to employer changes is mechanical because workers can only change industry if they also change employers.

Results in the previous paragraph are from a broad industry classification, but the magnitude of the between-major differences is very similar when a narrower classification of industry is used. When I categorize industries according to the employer's line of business within an economic sector (four-digit NAICS codes), rather than economic sectors (two-digit NAICS codes) as above, general majors change 20-30% more often than specific majors. This suggests that many of the narrow changes of industry are *between* economic sectors rather than *within*. That is, the changes of industry I measure are more likely to be changes between more distant lines of business – for example, between Securities & Commodity Exchanges and Elementary & Secondary Schools – than between similar lines of business – for example, Office Administrative Services, Employment Services & Business Support Service. Thus, not only do general majors switch industries more frequently, but their industry changes constitute an arguably more dramatic change of work context.

Fifth, the changes of general majors are more "complex" than those of specific majors: when a worker changes employers, the change is more likely to be across industries for general majors than for specific majors. For example, around three to four years after graduation 72% of the employer changes of general majors are also broad industry changes compared to 64% for specific majors, a difference that only marginally shrinks throughout the analysis window. These changes are more complex (Neal, 1999), as a change of only employer may correspond to a change in the management practices of the firm, whereas a change of employer and industry also corresponds to a change in the types of clients worked with or the tasks emphasize on the job. Thus, not only do general majors switch industries more frequently, but their industry changes constitute an arguably more dramatic change of work context.

Finally, as with earnings growth, job changing rates differ most dramatically between general and specific majors, and the outcomes for the twenty majors that are neither general nor specific fall between these two extremes. The difference in employer and industry changing between general majors and majors that are neither general nor specific are statistically significant at the 1% level but tend to not be economically substantive ($<$ 1 percentage point). The probability of switching occupations, however, does not significantly differ between these two groups but is of a larger magnitude. Panels C-D of Appendix Table A.12 also compares majors using a linear measure of occupational dispersion (percent of a major's recent graduates in the major's three largest occupations) rather than the three college major specificity groups. Majors with an occupation dispersion that is 10 percentage points above the mean (i.e. a major that is more specific than the average major) change employer industry and occupation less frequently than the average major.

1.6.1 Alternative Potential Mechanisms

General majors experience steeper earnings growth and change occupation, employer and industry more frequently than specific majors. This section investigates whether the average differences mask important heterogeneity in outcomes. In particular, I examine the role of major subject fields (e.g. Engineering or Education), gender differences, and graduate education.

Gender Differences Within gender, general majors change jobs more frequently and experience steeper earnings growth than specific majors. To illustrate this, Figure 1.5 plots estimates of Equation (1.3.1) run separately by gender. While the job changing differential between general and specific majors is sizable for both females and males, the magnitude is larger among females. More specifically, among females, general majors are 40-50% (5-7 percentage points) and 25% (5 percentage points) more likely to switch industry and employer than specific majors but among men the differential is 18% (4.5 percentage points) and 20-25% (3-5 percentage points), respectively. For both genders, specific majors initially earn more than general majors, but general majors shrink the earnings gap through higher earnings growth. However, among females, the initial earnings premium of specific majors is smaller, and the gap between majors closes more dramatically over time.

One reason why there may be gender differences in the *magnitude* of the general-specific major differences is that there are gender differences in the composition of major choice: the *typical* specific major among female college graduates is an Education or Health major, whereas it is a Computer Science & Engineering major among males (see Appendix Table A.6). As an illustration, I reweight the female observations to have an empirical distribution across all 61 detailed majors that is equivalent to that observed among males. Specifically, I run the female regression with weights equal to $p_{f,m}/p_{m,m}$ where $p_{f,m}$ ($p_{m,m}$) is the share of females (males) with each of the 61 majors. Figure 1.5d compares the specific majors' earnings premium for the unweighted and reweighted samples. When weights are applied, the magnitude of the general-specific earnings gap among females increases dramatically to a magnitude that is close to that observed among unweighted males. The results don't fully converge after reweighting, which is likely due to gender differences in outcomes within major subject fields (see Appendix Figure A.4).

Major Subject Field The average differences between general and specific major are not attributable to any one college major subject field (e.g. Health, Pure Sciences or Education), nor do they merely represent the contrast between technical or STEM (science, technology, engineering and math) and non-technical or non-STEM majors.

First, general and specific majors each contain 20 college majors that come from a variety of subject fields (see Appendix Table A.4). The 20 general majors include majors from the subject fields of Social Science, Humanities and Communications. While STEM majors including Computer Science & Engineering tend to be specific, specific majors also include non-STEM majors like Education and Health majors.

Second, while there are differences in job changing rates across the subject fields, the job changing rates of major subject fields that tend to be specific differ from those that tend to be general. Figure 1.6 illustrates this by plotting the results from the estimation of Equation (1.3.1) with indicators for nine major subject fields in place of the indicators for general and specific majors.[27] College graduates with Communications & Marketing, Humanities and Social Sciences majors, major subject fields which tend to be general, change jobs more frequently than Education, Health and Computer Science & Engineering majors, major subject fields which tend to be specific. In addition, with some exceptions, the more specific subject fields display lower earnings growth than the more general subject fields. The two exceptions are the subject fields of Pure Sciences and Computer Science & Engineering which display exceptionally steep earnings growth.

Third, the pattern of results within gender provide further evidence that general-specific major differences don't merely reflect the contrast between technical or STEM and non-technical or non-STEM majors. Among females, general-specific differences persist even though the typical specific major is not a STEM major but is Education or Health and

[27]The 61 detailed majors are divided into mutually exclusive and exhaustive categories including Business & Economics, Computer Science & Engineering, Communications & Marketing, Education, Health, Humanities, Pure Sciences and Social Sciences. The ninth category is "Other Majors" but is omitted from the figures and results discussion. See Appendix Table A.16 for mapping of the 61 majors into the 9 college major subject fields.

the typical general major is less likely to be Business & Economics compared to men.

Graduate Education In the initial 15 years post undergraduate degree, graduate degree attainment rapidly increases until almost 40% of bachelor's degree holders have some type of graduate degree. Appendix Figure A.5a plots graduate degree attainment levels in the public-use ACS by year since graduation for all 61 majors, and also the average among general and specific majors. The variation in attainment levels between all majors far exceeds the limited differences in mean attainment between general and specific majors. There are, however, more sizable differences in the *types* of degrees attained: specific majors are more likely to obtain Master's degrees and less likely to obtain Professional degrees than general majors (see Appendix Figure A.5b).[28]

A detailed analysis of the relationship between undergraduate major, graduate degree attainment and labor market outcomes is beyond the scope of this analysis, but I examine the sensitivity of general-specific major differences to various controls for graduate education.[29] I run a series of regressions that allows outcomes to evolve differently in the period prior to graduate enrollment, during enrollment and post graduate degree attainment. Specifically, I estimate a regression of the following form:

$$\begin{aligned} Y_{imt} = \beta_0 &+ \sum_{m \in M} \phi_m(\text{major group}_{im}) + \beta_g f(g_{it}) + \sum_{m \in M} \beta_{m,g}(f(g_{it}) \cdot \text{major group}_{im}) \\ &+ \sum_d \beta_{e,d}(\text{enroll}_{itd}) + \sum_d \beta_{a,d}(\text{attain}_{itd}) \\ &+ \sum_d \beta_{e,d,g}(\text{enroll slope}_{itd}) + \sum_d \beta_{a,d,g}(\text{attain slope}_{itd}) + \delta X_{it} + \theta_t + \gamma_t + \epsilon_{it} \qquad (1.6.1) \end{aligned}$$

where $f(g_{it})$ is quadratic and includes years since (undergraduate) graduation (g) and years since graduation squared (g^2), *major group*$_{im}$ are indicators for college major specificity group, and $f(g_{it}) \cdot$ *major group*$_{im}$ are interactions. The variables *enroll*$_{itd}$ and *attain*$_{itd}$ are graduate-degree type enrollment and attainment dummies equal to 1 if worker i is enrolled in or has attained graduate degree d in period t and 0 otherwise. Degree types (d) include Master's, Professional and Doctorate degrees. The term *enroll slope*$_{itd}$ is equal to 0 pre-enrollment, equal to 1 in the first year of enrollment and increases linearly until the worker has obtained a graduate degree after which point the term remains fixed. The attainment slope (*attain slope*$_{itd}$) is defined similarly but starts increasing from 1 in the first

[28] Graduate degree types are pre-defined in the ACS and NSCG. Master's degrees include MA, MS, MSW, MBA, MEd, Professional degrees include MD, DDS, DVM, LLB, JD and Doctorate degrees include PhD DSc and EdD.

[29] See Altonji and Zhong (2021) who provide a detailed analysis of the returns to a broad set of graduate degrees and the mapping between undergraduate major and graduate degrees.

year post graduate degree attainment. These functions are always zero for individuals that never enroll or attain graduate education. Finally, the vectors X_{it} and Z_i contain a full set of time-varying and time-invariant controls equivalent to those used in Equation (1.3.1).

I use the coefficients from Equation (1.6.1) to isolate the portion of the labor market trajectory that occurs prior to enrollment and attainment. For general majors, this is the quadratic profile mapped out by the coefficients β_g and for specific majors it is determined by β_g, ϕ_m and $\beta_{m,g}$.[30] If differences in graduate attainment are a significant contributor to between-major differences in earnings and job changing, then the pre-enrollment and attainment trajectory should differ substantially from the trajectory among all workers, which is estimated by Equation (1.6.1) without any graduate controls.

Results in Table 1.2 display point estimates from Equation (1.6.1) using the weighted and unweighted NSCG-LEHD, the analysis dataset which contains detailed information on the timing of graduation degree attainment. This is also demonstrated graphically in Appendix Figures A.6 and A.7, which plot the trajectories implied by the coefficients. For each outcome, Column (1) presents the baseline results without any graduate controls, Column (2) adds in enrollment and attainment dummies, and Column (3) adds in enrollment and attain slopes. Note that there are some differences in labor market trajectories (without graduate controls) between the ACS-LEHD sample – the primary analysis sample for all results so far – and the (un)weighted NSCG-LEHD – the analysis sample used here. See Appendix Section A.2.4 for a detailed discussion.[31]

For all outcomes, there are minimal difference in the coefficients across specifications. For both general and specific majors, including controls for graduate school (enrollment or attainment and dummies or slopes) does slightly decrease period-specific earnings growth, and to a larger extent in later years. This suggests that, unsurprisingly, graduate attainment is positively correlated with earnings growth. However, the evolution of the earnings gap is steady across specifications: the decline in the specific majors' earnings premium occurs among individuals prior to graduate enrollment and attainment. Results in Appendix Figure A.8 show that these conclusions also hold among workers in the public-use ACS. In particular, the earnings growth of both general and specific majors is lower when individuals who aren't enrolled or have graduate degrees are excluded, but the specific majors' earnings premium still decreases over time. Finally, the higher rates of employer and industry changing for

[30] To obtain trajectories for workers post graduate attainment, the coefficients $\beta_{a,d}$ and $\beta_{a,d,g}$ and are also needed.

[31] In general, across all three specifications general majors change employers and industries more often than specific majors throughout the early career. The evolution of the specific major's earnings premium differs somewhat: in the unweighted NSCG-LEHD sample it decreases steadily in the early career and is at a higher magnitude than in the ACS-LEHD. In the weighted NSCG-LEHD the differential follows a u-shape pattern, because the gap between general and specific majors increases substantially 13-15 years post graduation.

general relative to specific majors are also robust, albeit slightly smaller, when graduate controls are included.

1.7 Earnings Growth Accounted for by Job Changing

To what extent are mean differences in earnings growth between general and specific majors accounted for by mean differences in the frequency of job changes? For each years since graduation bin (g), I use an Oaxaca-Blinder decomposition to divide the earnings-growth difference between general and specific majors into a part that is "explained" by mean differences in job changing rates, a part "explained" by group differences in other covariates, and a part "unexplained" by mean differences in all covariates (Oaxaca, 1973; Blinder, 1973).

Specifically, the key quantity to be decomposed in each bin g is the difference in mean earnings growth between general majors (*gen*) and specific majors (*spec*):

$$E[\Delta earn_{i,gen}] - E[\Delta earn_{i,spec}] \tag{1.7.1}$$

where $E[\Delta earn_{i,m}]$ denotes the expected value for major m of individual-level changes in log annual earnings, defined as $\Delta(earn_{i,m,t}) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$. The subscript g is dropped for notational ease. Following the standard Oaxaca-Blinder approach, Equation (1.7.1) can be expressed as the difference of major-specific regression models evaluated at the mean, defined as $E[\Delta earn_{i,m}] = E[\sum_k \beta_{k,m} X_{k,i} + u_{i,m}]$ for each major m. Let $\hat{\beta}_{k,m}$ be the least-squares estimate of $\beta_{k,m}$ for each variable X_k and the sample mean $\overline{\Delta earn_{i,m}}$ be an estimate for $E[\Delta earn_{i,m}]$. Then substituting these terms into Equation (1.7.1), and rearranging, adding and subtracting terms yields:

$$\begin{aligned}\overline{\Delta earn_{i,gen}} - \overline{\Delta earn_{i,spec}} = &\sum_k \hat{\beta}_{k,P}(\overline{X}_{k,gen} - \overline{X}_{k,spec}) \\ &+ \sum_k \overline{X}_{k,gen}(\hat{\beta}_{k,gen} - \hat{\beta}_{k,P}) + \sum_k \overline{X}_{k,spec}(\hat{\beta}_{k,P} - \hat{\beta}_{k,spec})\end{aligned} \tag{1.7.2}$$

where $\overline{X}_{k,m}$ is the mean of X_k among major m, and $\hat{\beta}_{k,P}$ is the least-squares estimate for $\beta_{k,P}$ from a pooled model discussed below. The first term in Equation (1.7.2) is commonly referred to as the *explained gap* and the remaining terms as the *unexplained gap*. The explained gap measures the expected change in the earnings-growth difference (in a statistical, not causal, sense) if specific majors are assumed to have the same means of the covariates as general majors. It can be further subdivided into parts accounted for by each individual covariate X_k or by groups of covariates.

The explained gap in Equation (1.7.2) weights mean differences in the covariates using the coefficients from a pooled model ($\hat{\beta}_{k,P}$). The pooled regression is estimated using observations for all workers and constrains coefficients to be the same for all workers irrespective of college major.[32] As the pooled model includes all covariates used in the decomposition, the coefficients $\hat{\beta}_{k,P}$ don't suffer from path dependence, a situation in which the decomposition results depend on the order that covariates are added to the model (Gelbach, 2016). Finally, the model also includes dummies for college major specificity group as an additional covariate to prevent residual group differences from spilling over into the parameters of the pooled model (Jann, 2008; Elder et al., 2010; Fortin et al., 2011).[33]

Why use coefficients from a pooled model? It is well understood that the results of the Oaxaca-Blinder decomposition depend on the reference parameters used to calculate the explained outcome gap. There is usually no single correct choice of reference parameter, and in this case, the use of parameters from a pooled model implicitly assumes away between-major differences in the earnings growth premium of job changers over job stayers. Analysis in Appendix A.1.1 investigates this assumption.

1.7.0.1 Results

Results are presented in Figure 1.7 and Appendix Table A.13 for the ACS-LEHD sample. Average differences in the covariates between general and specific majors generally account for around half of the raw earnings-growth difference. To illustrate this, Figure 1.7a plots three things: (1) the unadjusted (raw) earnings-growth difference between general and specific majors, (2) the portion of the earnings-growth difference (in log earnings) explained by all of the covariates, and (3) the portion of the earnings-growth difference (in log earnings) unexplained. In all periods except 1-2 years post graduation, the raw earnings-growth difference is positive and ranges from .5-2 percentage points because the average of individual-level earnings changes among general majors exceeds that of specific majors. During these same periods, the explained earnings-growth difference is positive and ranges from 45-60%: adjusting for mean differences in all covariates *reduces* the

[32] This contrasts with the traditional form of the decomposition in which the explained portion is $\sum_k \hat{\beta}_{k,spec}(\overline{X}_{k,gen} - \overline{X}_{k,spec})$ and the unexplained portion is $\sum_k \overline{X}_{k,gen}(\hat{\beta}_{k,gen} - \hat{\beta}_{k,spec})$. Mean differences in the covariates are weighted by the coefficients for specific majors ($\hat{\beta}_{k,spec}$) rather than by the coefficients from a pooled model ($\hat{\beta}_{k,P}$). Equation (1.7.2) is yielded by adding and subtracting terms $\sum \hat{\beta}_{k,P}\overline{X}_{k,spec}$ and $\sum \hat{\beta}_{k,P}\overline{X}_{k,gen}$.

[33] Even though the decomposition only concerns the earnings-growth difference between general and specific majors, the sample used to estimate the pooled model also includes workers with majors that are neither general nor specific. I use the Jann (2008) stata "oaxaca" procedure and the "pooled" option. I first estimate the pooled model and then feed in the pooled model coefficients to the decomposition using the "reference()" option.

earnings-growth difference between majors. The unexplained portion is generally positive, as including all of the covariates doesn't drive the earnings-growth different to zero.

During most years since graduation bins, mean differences in the frequency of job changes accounts for the largest share of the earnings-growth difference explained by the observable covariates. Figure 1.7b shows this by plotting the percent of the explained earnings-growth difference accounted for by each of the covariate groups. Job changing alone accounts for 30-50% of the explained earnings-growth difference. The equivalent shares for demographic and economic covariates are generally lower, and range from 15-40% and 22-37%, respectively. Finally, as a percent of the *unadjusted (raw)* earnings-growth difference, job changing alone accounts for 15-27%.[34]

The results of this decomposition exercise suggest that differences between general and specific majors in the frequency of job changes, can account for a non-trivial portion of the earnings-growth difference. In particular, adding in control variables for job changing, reduces the regression-estimated earnings growth advantage of general majors. As a large portion of the earnings-growth difference between general and specific majors remains unaccounted for, there are other salient (un)observable factors that may further explain earnings growth differences.

1.8 Conclusion

While previous empirical evidence suggests substantial earnings premiums associated with four-year college enrollment and completion (Goldin and Katz, 2009), the labor market outcomes of college graduates are far from uniform. In addition, differences in average earnings between college majors rival the mean earnings gap between high school and college graduates (Altonji et al., 2012, 2016a). Less well-understood is the extent to which mean earnings differences between majors evolve over the career due to cross-major heterogeneity in earnings growth. Moreover, there is less direct evidence on what characteristics of college majors explain differences in earnings growth, or what mechanisms contribute to these differences.

In this paper, I contrast the labor market outcomes of graduates with skills that are more tailored to particular occupations (specific majors) and graduates with less tailored skills (general majors). I focus on this dimension of heterogeneity, as prior work emphasizes a contrast between more general and specific education in labor market outcomes. In a simple

[34]The major exception to this general pattern of results is the period spanning 1-2 years post graduation. During this period, the earnings growth of specific majors exceeds that of general majors by 3.4 percentage points. This is similar to the estimate of a 2.9 percentage point differential from the earnings change regression (see Appendix Table A.14 and Appendix Figure A.1). As the explained earnings-growth difference is negative, adjusting for mean differences in the covariates *widens* the earnings-growth difference between majors.

conceptual framework, I illustrate how differences between general and specific majors in the link between skills and occupations can mediate earnings growth through the mechanism of job search. I create the first-ever linkage of two nationally representative surveys with a large administrative earnings dataset to provide novel descriptive evidence on the job changing patterns of general and specific majors.

I find that the earnings gap of general and specific majors evolves over time as the two groups experience different rates of earnings growth. Specific majors initially earn more than general majors but general majors close the earnings gap within thirteen years post graduation. Consistent with prior work on job changing, I find that early in the career college graduates change jobs frequently and but do so less frequently with time in the labor market. I then provide new evidence that these patterns differ between college graduates with general and specific majors: general majors are substantively more likely to change employer, industry and occupation than specific majors. Differences exists for both narrow and broad industry and occupation categorizations. The changes of general majors are also more "complex" than those of specific majors (Neal, 1999), as their employer changes are more likely to also be industry changes. Extensive variation in job changing and earnings growth is also present across a wide range of major subject fields (e.g. Education, Engineering, Health or Humanities). General-specific major labor market differences exist with gender and are not substantially affected by accounting for graduate attainment.

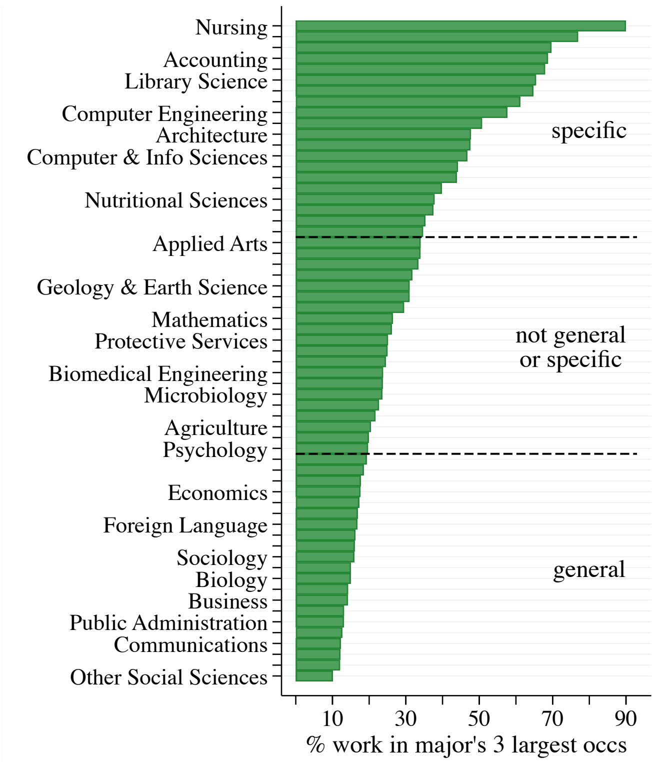
The results in this paper indicate that post-secondary investments based on labor market outcomes should be carefully made. Earnings and employment outcomes by institution and major are increasingly accessible in publicly available online tools, and this information can be used by students when selecting college major. Moreover, federal- and state-level policymakers are increasingly advocating for the use of labor market outcomes in the funding formula that allocate post-secondary funding across institutions and programs (Minaya and Scott-Clayton, 2019). The estimates of this paper, however, indicate that the relative labor market outcomes of college majors differ between early- and mid-career. As a result, the optimal post-secondary investments for both students and policymakers will depend on *which* outcomes are used and in particular *when* the outcomes are measured. The estimates of this paper are descriptive and do not reflect the pure causal impact of college major choice on labor market outcomes. Future work could explore the extent to which these differences reflect the causal impact of choosing a general versus specific major. A fruitful approach could be the use of quasi-experimental variation in student assignment to different majors (Hastings et al., 2013; Kirkeboen et al., 2016).

Table 1.1: Individual-Level Summary Statistics for the Analysis Samples

sample	Public-use ACS			ACS-LEHD			NSCG-LEHD (weighted)		
college major group:	all	general	specific	all	general	specific	all	general	specific
demographics (%):									
female	56.19	54.81	58.40	59.53	57.47	62.95	59.13	57.89	57.58
black	7.51	7.90	6.97	4.83	5.23	4.29	8.70	9.65	7.58
hispanic	6.11	6.21	5.83	7.44	8.05	5.98	9.57	12.28	7.58
college major specificity (%):									
general	44.91	-	-	45.43	-	-	49.57	-	-
specific	29.64	-	-	29.24	-	-	28.70	-	-
not general or specific	25.45	-	-	25.20	-	-	22.61	-	-
curriculum-based major groups (%):									
Business & Economics	19.35	29.66	10.29	19.84	29.9	10.71	22.61	33.33	D
Computer Sci/Engineer	12.40	-	35.86	9.14	-	27.68	9.57	-	30.30
Comm/Marketing	8.36	12.48	-	9.40	14.08	-	8.26	11.40	-
Education	8.91	-	30.07	10.31	-	35.27	8.70	-	30.30
Health	6.28	1.37	19.12	6.79	1.49	20.98	6.96	2.63	19.70
Humanities	14.01	21.46	-	13.84	21.55	-	16.52	26.32	-
Other Majors	4.75	5.07	0.08	5.09	5.06	0.09	4.35	4.39	D
Pure Sciences	11.35	14.72	0.81	9.66	12.36	0.80	8.26	9.65	0.45
Social Sciences	14.58	15.24	3.77	15.80	15.52	4.46	16.52	14.04	4.55
observations (N):									
individuals	1,376,715	616,317	409,187	383,000	174,000	112,000	11,500	4,400	4,100
person x year	1,376,715	616,317	409,187	2,637,000	1,172,000	648,000	76,000	37,000	22,000

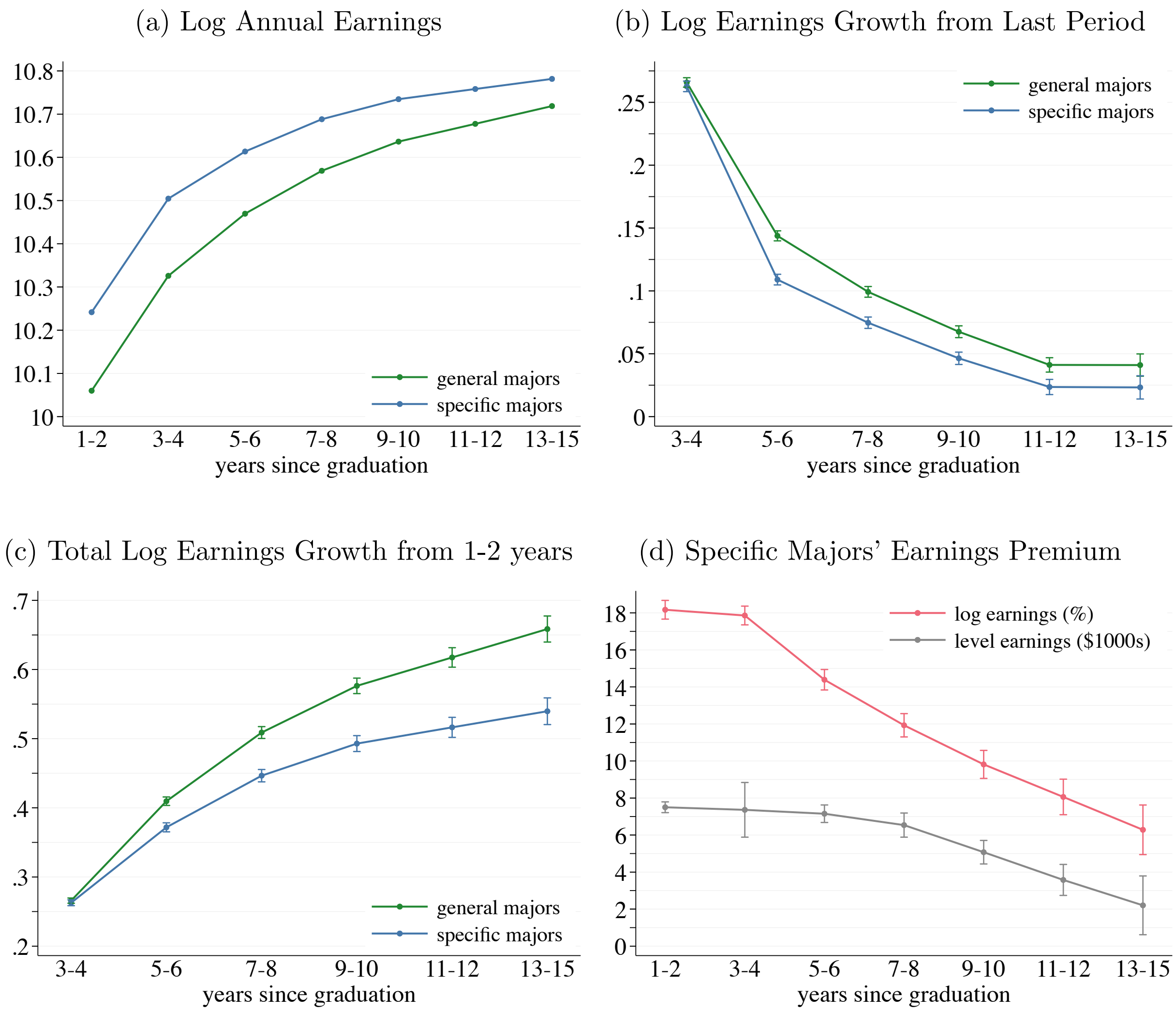
Note: Table displays summary statistics for all individuals from three different samples: (1) the public-use 2009-2019 ACS, (2) the ACS-LEHD analysis sample, and (3) the NSCG-LEHD analysis sample. Column (1) consists of all employed four-year college graduates, that are from the 1999-2012 graduating cohorts, 1-15 years post graduation. Columns (2) and (3) include four-year college graduates from the 1999-2012 graduating cohorts with 3+ quarters of non-zero LEHD earnings in the first year post graduation in one of the 23 covered states (see Appendix Figure A.10). Column (1) weighted using ACS survey weights and Column (3) weighted using NSCG survey sample weights. College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). Results were disclosed by the U.S Census Bureau's Disclosure Review Board with approval numbers. All cell counts are rounded. "D" indicates cells that have been deleted during disclosure review.

Figure 1.1: Employment Share in Major's 3 Largest Occupations



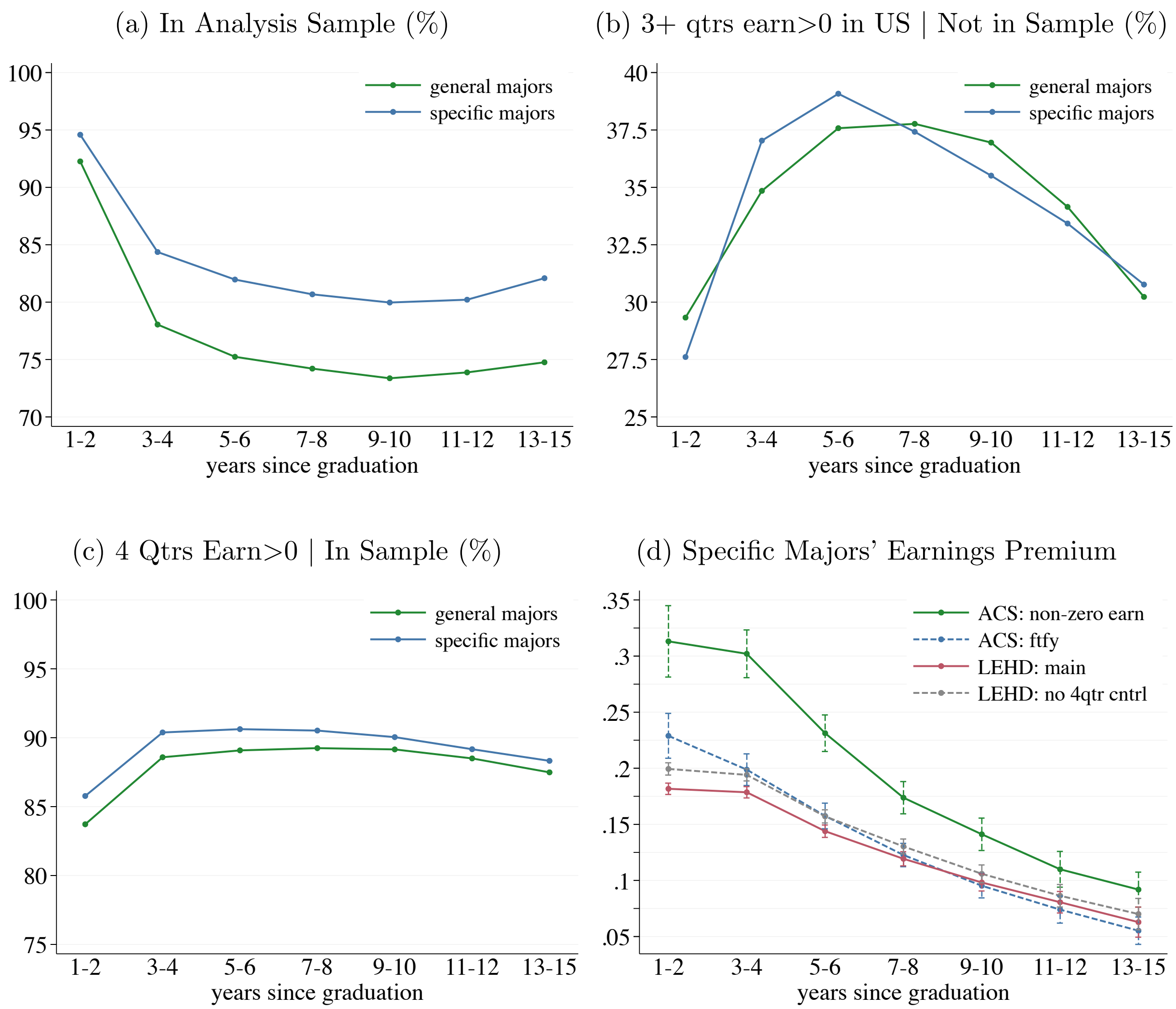
Note: Figure displays the percent of each major's total employment accounted for by the three largest occupations for the major. Data source is the public-use 2009-2019 ACS. Workers are restricted to unenrolled college graduates who are employed 1-5 years post undergraduate degree and that work at least part-time part-year (at least 27-39 weeks/year and 20 hours/week). The 20 majors with graduates that are clustered in a small number of occupations (i.e. a high percent of graduates employed in the major's three largest occupations) are considered specific majors. General majors are the 20 majors with the widest dispersion of graduates across occupations. All other majors are not general or specific. Appendix Table A.2 contains the results for all majors.

Figure 1.2: Earnings Differences between General & Specific Majors



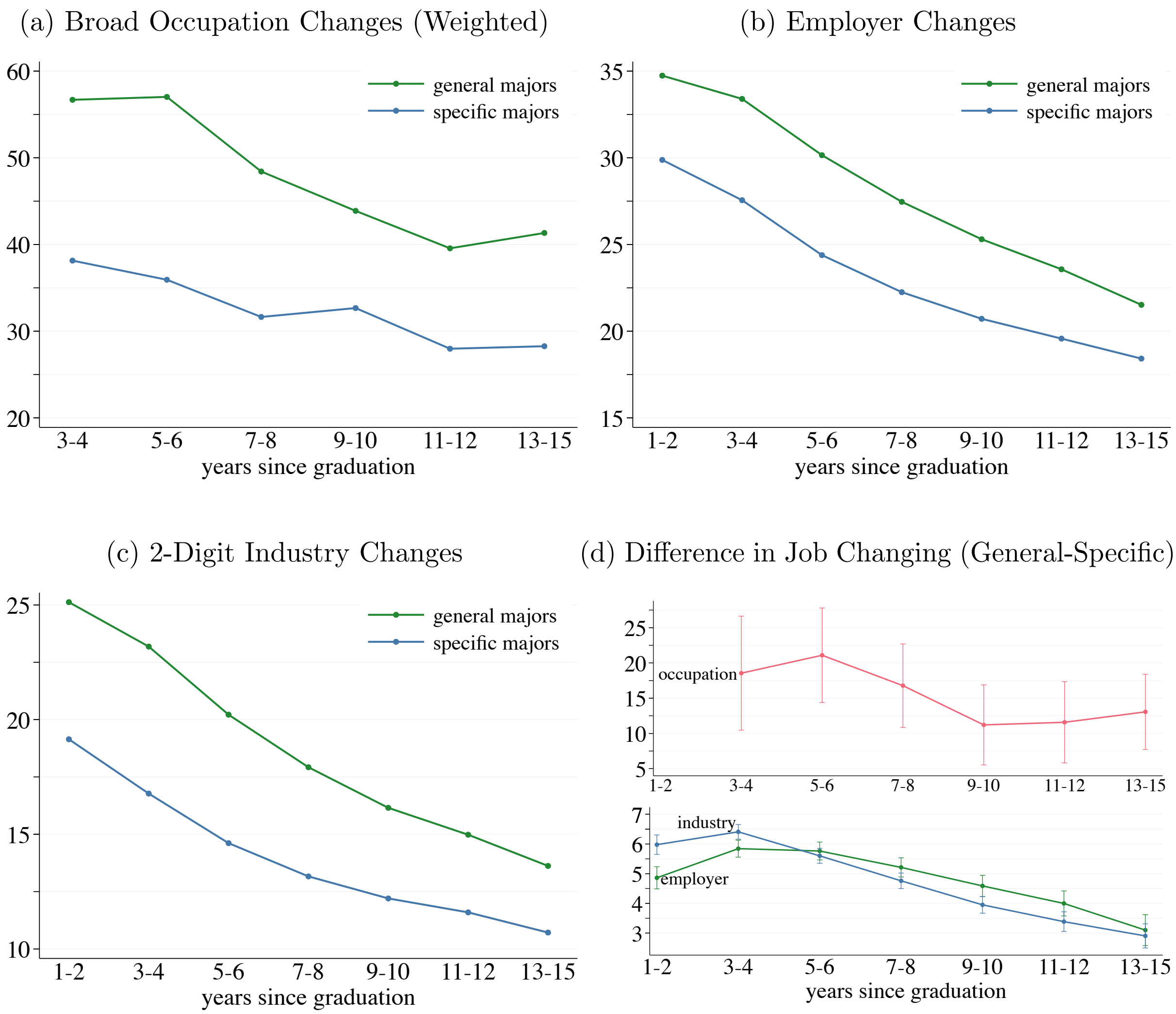
Note: Figure displays estimates of (a) log annual earnings, (b) earnings growth from the previous period, (c) cumulative earnings growth and (d) the earnings premium of specific majors over general majors. Plotted are the coefficients from a regression of (log) annual earnings on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). Earnings levels for general majors in bin g are calculated as $\beta_0 + \beta_g$ where β_0 is the sample mean of the outcome among general majors in the omitted period and for specific majors is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. Earnings growth is calculated using changes in mean earnings between major x years since graduation cells: equal to $(\beta_g - \beta_{g-1})$ from period $g-1$ to g for general majors and $(\beta_{m,g} - \beta_{m,g-1}) + (\beta_g - \beta_{g-1})$ for specific majors. Cumulative earnings growth from 1-2 years since graduation to period g is β_g for general majors and $\beta_g + \beta_{m,g}$ for specific majors. The specific major earnings premium is $\beta_{m,g} + \phi_m$. Data source is the ACS-LEHD, see Section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Standard errors are clustered at the individual level. Regression includes 2,637,000 observations for 383,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. See Appendix Table A.7 and A.8 for regression results.

Figure 1.3: Differences in Employment between General & Specific Majors



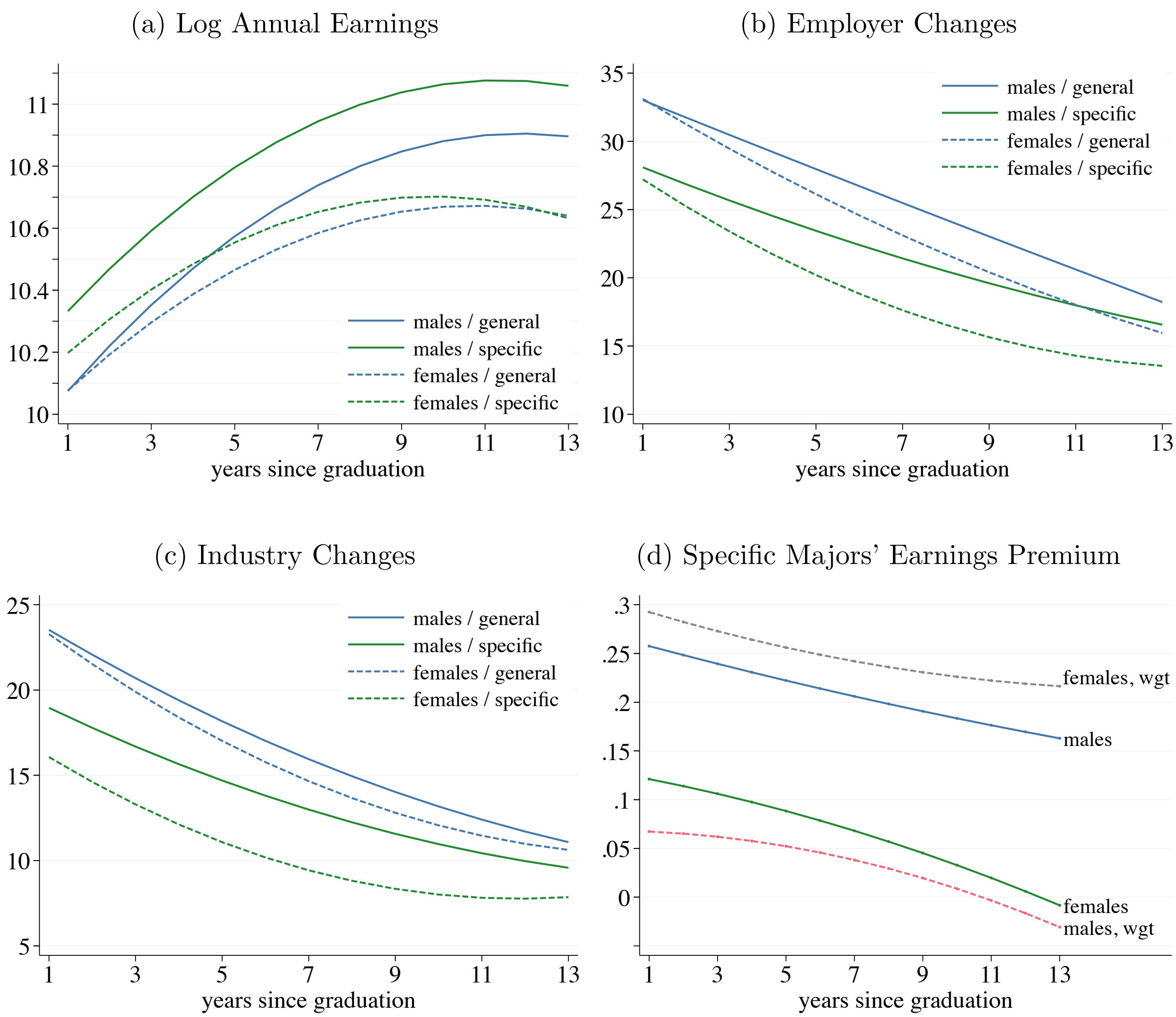
Note: Figure displays estimates of (a) the probability of being in the ACS-LEHD analysis sample (3+ quarters of earn>0), (b) the probability of 3+ quarters of earn>0 in LEHD nationally conditional on not being in the ACS-LEHD analysis sample (c) the probability of having 4 rather than 3 quarters earn>0 (conditional on being in the analysis sample), and (d) the specific majors' log earnings premium for various samples. Coefficients are from a regression of the outcome on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). In Panel (a) and (b) the outcome for general majors in bin g is $\beta_0 + \beta_g$ where β_0 is the sample mean among general majors in the omitted period and for specific majors is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. In Panel (c) estimates are P(3+ quarters earn>0 nationally)-P(3+ quarters earn>0 in 23 states) divided by 1-P(3+ quarters earn>0 in 23 states), where P(3+ quarters earn>0 in 23 states) are the estimates in panel (a) and P(3+ quarters earn>0 nationally) are calculated using the LEHD U.S. indicators files. In Panel (d) the specific major earnings premium is $\beta_{m,g} + \phi_m$. Panel (a) and (b) include all observations for all individuals in the ACS-LEHD, irrespective of number of quarters with earn>0 (3,396,000 observations for 383,000 individuals). Panel (c) only includes observations with 3+ quarters earn>0 (2,637,000 observations for 383,000 individuals). In Panel (d) "LEHD: main" is as depicted in Figure 1.2d and "LEHD:no-4qtr cntrl" is the earnings gap when the dummy for 3 rather than 4 quarters of non-zero LEHD earnings is excluded. Estimates for "ACS" are from the 2009-2019 public-use ACS and includes all employed four-year college graduates, that are from the 1999-2012 graduating cohorts, 1-15 years post graduation. "Ftfy" is defined as 35+hours/week and 40+weeks/year. All regression include controls as described in Section 1.3. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board.

Figure 1.4: Job Changing Rates for General & Specific Majors



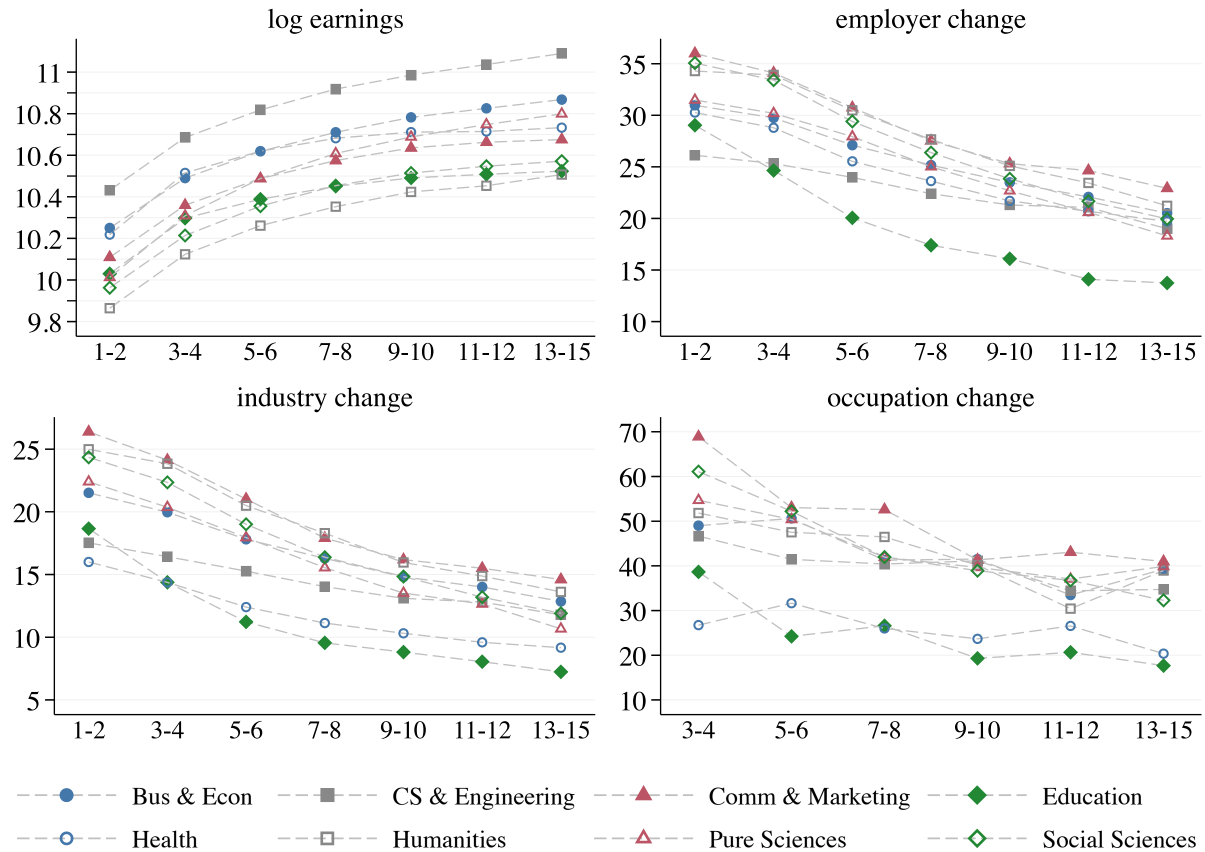
Note: Figure displays estimates of (a) broad occupation change, (b) employer change, (c) industry change (2-digit NAICS) and (d) the differential in each job changing outcome between general and specific majors. Coefficients are from a regression of the outcome on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). Employer and industry changes are defined by change in a worker's main job employer (employer at which they earned the highest earnings) from the previous year. Occupation changes is defined by change in a worker's occupation from two years prior. There are 20 broad occupations categories. In Panels (a)-(c) the outcome for general majors in bin g is $\beta_0 + \beta_g$ where β_0 is the sample mean among general majors in the omitted period and for specific majors is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. In Panel (d) the difference is calculated as is -$phi_m + \beta_{m,g}$. Data source for Panel (b) and (c) is the ACS-LEHD (2,254,000 observations for 360,000 individuals) and for Panel (a) is the ACS-NSCG (61,500 observations for 35,000 individuals), see Section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Panel (d) estimates are weighted using NSCG survey weights. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. See Appendix Table A.9, A.10 and A.11 for regression results.

Figure 1.5: Gender Differences in the General-Specific Gap



Note: Figure displays estimates of (a) log annual earnings, (b) employer change, (c) industry change (2-digit NAICS) and (d) the specific majors' earnings premium for general and specific majors and separately by gender. Coefficients are from gender-specific regressions of the outcome on a quadratic function of years since graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). Employer and industry changes are defined by change in a worker's main job employer (employer at which they earned the highest earnings) from the previous year. Occupation changes is defined by change in a worker's occupation from two years prior. There are 20 broad occupations categories. In Panels (a)-(c) the outcome for female (male) general majors in bin g is $\beta_0 + g\beta_g + g^2\beta_{g2}$ where β_0 is the sample mean among female (male) general majors in the omitted period and for female (male) specific majors is $\beta_0 + \phi_m + g(\beta_g + \beta_{g,m}) + g^2(\beta_{g2} + \beta_{g2,m})$. In Panel (d) the female (male) specific majors' earnings premium is $g\beta_{m,g} + g^2\beta_{m,g} + \phi_m$. "Wgt" refers to outcomes from a weighted regression. The weights are calculated using share of females (males) with major m: $p_{f,m}$ ($p_{m,m}$). Outcomes labeled "males, wgt" are male observations reweighted by $p_{m,m}/p_{f,m}$. Data source for all panels is the ACS-LEHD. In Panel (a) and (d) 1,557,000 female and 1,080,000 male observations. In Panel (b) and (c) 1,329,000 female and 926,000 male observations. Section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

Figure 1.6: Earnings Growth and Job Changing by College Major Subject Field



Note: Figure displays estimates of (a) log annual earnings, (b) employer change, (c) industry change (2-digit NAICS) and (d) occupation change separately for 8 college major subject fields. Coefficients are from regressions of the on two-year years since graduation bins, college major subject field dummies and interactions of the two (similar to Equation (1.3.1)). Majors are grouped into nine mutually exclusive college major subject fields and Education is the omitted major (see Appendix Table A.16). Estimates for "All Other Majors" are omitted from the graph. Employer and industry changes are defined by change in a worker's main job employer (employer at which they earned the highest earnings) from the previous year. Occupation changes is defined by change in a worker's occupation from two years prior. In Panels (a)-(c) the outcome for Education majors in bin g is $\beta_0 + \beta_g$ where β_0 is the sample mean among Education majors in the omitted period and for subject field m is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. Data source in Panel (a)-(c) is the ACS-LEHD: (a) 2,637,000 observations for 383,000 individuals, (b) and (c): 2,254,000 observations for 360,000 individuals. Data source in Panel (d) is the ACS-NSCG: 61,500 observations for 35,000 individuals. Panel (d) estimates are weighted using NSCG survey weights. See section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

Table 1.2: Graduate Education, Earnings Growth and Job Changing

Panel A: Weighted

	log earnings			employer change			industry change		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
	none	degree dummy	degree slopes	none	degree dummy	degree slopes	none	degree dummy	degree slopes
general majors:									
intercept	10.15	10.18	10.18	.3320	.3265	.3265	.2457	.2440	.2440
years grad	.1347***	.1329***	.1331***	-.0346***	-.0359***	-.0360***	-.0342***	-.0351***	-.0352***
	(.0137)	(.0135)	(.0135)	(.0082)	(.0082)	(.0082)	(.0066)	(.0066)	(.0066)
(years grad)^2	-.0064***	-.0066***	-.0067***	.0018**	.0019***	.0020***	.0019***	.0020***	.0021***
	(.0009)	(.0009)	(.0009)	(.0007)	(.0007)	(.0007)	(.0006)	(.0006)	(.0006)
specific majors:									
intercept	.2997***	.3068***	.3002***	-.0677***	-.0693***	-.0693***	-.1087***	-.1094***	-.1093***
	(.0363)	(.0363)	(.0364)	(.0205)	(.0204)	(.0204)	(.0175)	(.0176)	(.0176)
years grad	-.0450***	-.0465***	-.0461***	.0082	.0085	.0087	.0223***	.0225***	.0227***
	(.0135)	(.0132)	(.0132)	(.0092)	(.0091)	(.0092)	(.0073)	(.0073)	(.0073)
(years grad)^2	.0030***	.0029***	.0030***	-.0007	-.0006	-.0006	-.0016**	-.0015**	-.0015**
	(.0011)	(.0011)	(.0011)	(.0008)	(.0008)	(.0008)	(.0006)	(.0006)	(.0006)
constant	10.65***	10.67***	10.66***	.3270***	.2397***	.2411***	.2304***	.1987***	.2004***
	(.1285)	(.1285)	(.1286)	(.0643)	(.0736)	(.0736)	(.0497)	(.0556)	(.0556)

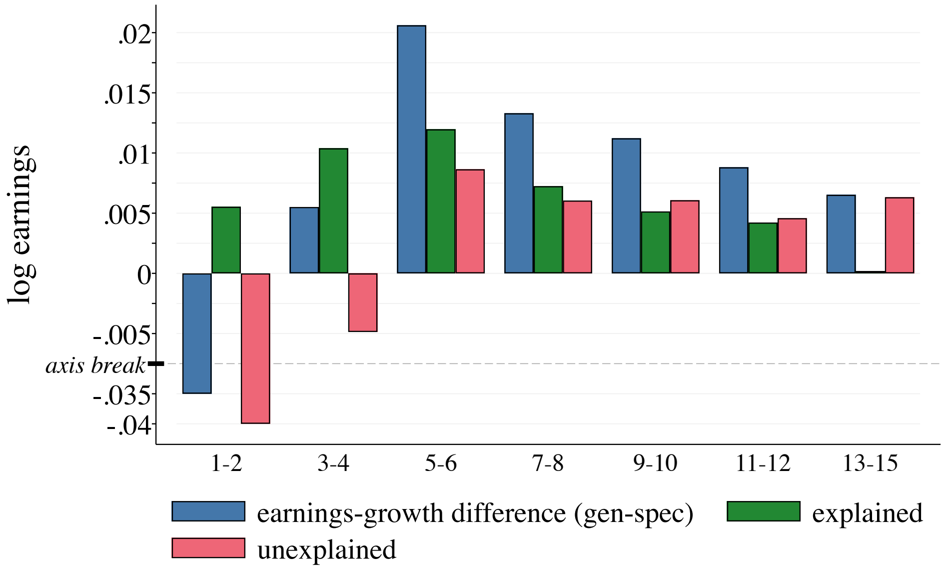
Panel B: Unweighted

	log earnings			employer change			industry change		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
	none	degree dummy	degree slopes	none	degree dummy	degree slopes	none	degree dummy	degree slopes
general majors:									
intercept	10.15	10.18	10.18	.3320	.3265	.3265	.2457	.2440	.2440
years grad	.1347***	.1329***	.1331***	-.0346***	-.0359***	-.0360***	-.0342***	-.0351***	-.0352***
	(.0137)	(.0135)	(.0135)	(.0082)	(.0082)	(.0082)	(.0066)	(.0066)	(.0066)
(years grad)^2	-.0064***	-.0066***	-.0067***	.0018**	.0019***	.0020***	.0019***	.0020***	.0021***
	(.0009)	(.0009)	(.0009)	(.0007)	(.0007)	(.0007)	(.0006)	(.0006)	(.0006)
specific majors:									
intercept	.2997***	.3068***	.3002***	-.0677***	-.0693***	-.0693***	-.1087***	-.1094***	-.1093***
	(.0363)	(.0363)	(.0364)	(.0205)	(.0204)	(.0204)	(.0175)	(.0176)	(.0176)
years grad	-.0450***	-.0465***	-.0461***	.0082	.0085	.0087	.0223***	.0225***	.0227***
	(.0135)	(.0132)	(.0132)	(.0092)	(.0091)	(.0092)	(.0073)	(.0073)	(.0073)
(years grad)^2	.0030***	.0029***	.0030***	-.0007	-.0006	-.0006	-.0016**	-.0015**	-.0015**
	(.0011)	(.0011)	(.0011)	(.0008)	(.0008)	(.0008)	(.0006)	(.0006)	(.0006)
constant	10.65***	10.67***	10.66***	.3270***	.2397***	.2411***	.2304***	.1987***	.2004***
	(.1285)	(.1285)	(.1286)	(.0643)	(.0736)	(.0736)	(.0497)	(.0556)	(.0556)
N	72000	72000	72000	60500	60500	60500	60500	60500	60500

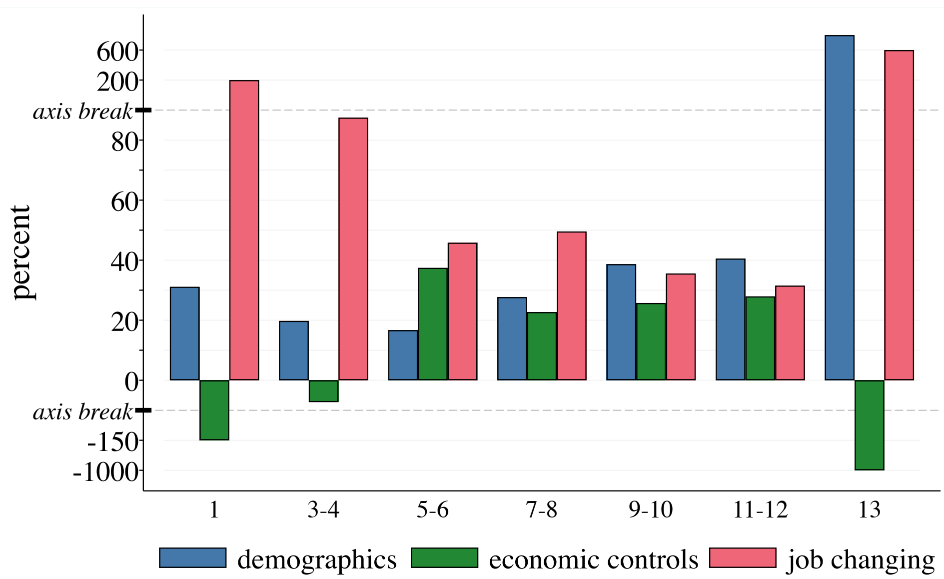
Note: Table displays estimates of (a) log annual earnings, (b) employer change, (c) industry change (2-digit NAICS) from regressions with controls for graduate education. Each column is a separate regression. For each outcome, estimates in Column (1) are from a regression of the outcome on a quadratic in years since graduation, college major specificity group, interactions of the two. Column (2) adds in a total of 6 graduate degree-type indicators for enrollment and attainment (Master's, Professional, Doctorate): $enroll_{itd}$ and $attain_{itd}$. Column (3) adds in the enrollment and attainment slopes: $enroll\ slope_{itd}$ and $attain\ slope_{itd}$. See Equation 1.6.1. For general majors "years grad" is equal to β_g and (years grad)^2 is equal to β_{g2}. For specific majors there are equal to $\beta_{m,g}$ and $\beta_{m,g2}$ respectively. The intercept for specific majors is ϕ_m and for general majors is the sample mean for general majors in the omitted period. Employer and industry changes are defined by change in a worker's main job employer from the previous year. All regression include controls as described in Section 1.3. Data source is the NSCG-LEHD. See section 1.2.2 for details on the analysis sample. Panel (a) estimates are weighted using NSCG sample weights. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. The implied trajectories are plotted in Appendix Figures A.6 and A.7. *p< 0.1, **p<0.05, ***p<0.01.

Figure 1.7: Accounting for Earnings Growth Differences between Majors

(a) Raw Earnings-Growth Difference: Explained and Unexplained by Covariates



(b) Percent of Explained Earnings-Growth Difference Explained by....



Note: Figure plots (a) the raw earnings-growth difference between general and specific majors and the portion explained and unexplained by covariates in a Oaxaca-blinder style decomposition. Figure (b) plots the percent of the explain gap that is attribute to each covariate group. A separate Oaxaca-blinder style decompositions is performed in each year since graduation bin. See Section 1.7 for details. The outcome is individual-level earnings growth: $\Delta(log(earn_{imt})) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$. In Panel (a) unadjusted (raw) earnings-growth difference is equivalent to $\overline{\Delta(log(earn_i))}_{gen}$-$\overline{\Delta(log(earn_i))}_{spec}$. In Panel (b) for each covariate, the figure plots the share of the total explained earnings-growth difference accounted for by only that covariate X_j: $\big(\beta_{j,P}(\overline{X}_{j,gen} - \overline{X}_{j,spec}\big)/\big(\sum_k \beta_{k,P}(\overline{X}_{k,gen} - \overline{X}_{k,spec})\big)$, where the denominator is the total explained earnings-growth difference. The covariate groups include (1) demographics (female, black, hispanic, and five-year bins for college graduation cohort), (2) economic controls (de-meaned unemployment rate at graduation, state and year fixed effects and a part-year employment indicator), (3) job changing (employer and industry changes). Data source is the ACS-LEHD. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. See Appendix Table A.13 for full results.

CHAPTER II

College Majors and Skills: Evidence from the Universe of Online Job Ads with Brad Hershbein, Steve Hemelt and Kevin Stange

2.1 Introduction

The choice of college major is one of the most direct ways for college graduates to acquire skills and signal competencies to employers. Indeed, earnings differences among college graduates with different majors can be larger than earnings differences between college and high school graduates (Altonji et al., 2012; Webber, 2014). Some of the earnings heterogeneity among majors is undoubtedly due to selection, but recent evidence also points to the importance of human capital development from the major itself (Hastings et al., 2013; Kirkeboen et al., 2016). College major provides much of the structure for the courses students take and thus the competencies and skills they develop during college. Because demand for certain skills has grown in recent years (Deming, 2017; Atalay et al., 2020), it is possible that employers' perceptions of the skills associated with graduates from different majors plays a large role in explaining earnings heterogeneity among college graduates. Somewhat surprisingly, however, there is little work that systematically characterizes the skills employers associate with college majors and their relation to differences in earnings.[1]

To start to fill this void, this paper answers two main questions: First, how does employer skill demand differ across majors? For example, is the desire for social skills concentrated among job postings in only a few majors or is it widely demanded across majors? Second, how does skill variation relate to earnings variation across majors? In answering these questions,

[1]In contrast, recent research has documented the importance of skill heterogeneity between and within occupations in explaining spatial wage variation (Deming and Kahn, 2018). But because occupation reflects post-labor-market selection, the role of pre-market skill acquisition as captured by college major remains underexplored.

we develop a new measure of the specificity of college majors based on their patterns of skill concentration. We also explore the role of place as it relates to within-major, cross-area differences in skill demand and earnings.

We measure the skills employers associate with particular majors using job vacancy data obtained from Burning Glass Technologies (BGT), comprising the near universe of all job ads from 2010-2018.[2] A unique feature of this data source–beyond its scale and universality–is the inclusion of information on majors, detailed skills, locations, and occupations, which permits us to characterize demand along these dimensions. In contrast to previous studies that document skill-major linkages mediated through occupation (Altonji et al., 2016b; Long et al., 2015), the job postings data allow us to measure skill-major linkages at the individual job level and to account for substantial within-occupation variation in skill demand (which may be correlated with college major). Moreover, this information precedes the employment choices of individuals, and is thus a more proximate and direct signal of skill demand independent of occupational sorting.

To answer our descriptive questions we take advantage of the more than 15,000 unique and detailed skills listed in job ads to create a tractable number of skill composites, adapting the approach of Deming and Kahn (2018). With these composites, we construct skill location quotient indices by major, similar to the approach typically used to measure industrial or occupational concentration. More specifically, we compare the vector of skills listed among job ads for each major to the vector of skills among jobs ads for all college-educated workers. The relative over- or under-representation of certain skills within a major provides evidence on the specificity of that field of study. We then construct major-specific skill vectors for each metropolitan statistical area (MSA). This permits us to examine the extent to which variation across MSAs in major-specific earnings can be explained by functions of their granular skill differences.

Our analysis reveals marked differences in the skills associated with different majors. Some skills–even composites–are concentrated within a small subset of majors whereas others are near universal. Employers demand social and organizational skills at similar rates across all majors, but customer service and financial skills appear specialized to relatively few majors. In turn, we find some majors are more typical of overall skill demand than others. For example, average skill demand for Business, Economics, and General Engineering majors accords reasonably closely with the average skill demand across all majors. Nursing, Education, and Foreign Language, on the other hand, are more specific, with jobs ads requesting skills demanded relatively infrequently in other majors. Together these results

[2] In 2021, after we acquired the data, Burning Glass Technologies merged with EMSI, a similar firm, and the company is now known as EMSI Burning Glass.

suggest that employers view majors as meaningfully encompassing different skill bundles.

Further evidence that employers view majors as a bundle of skills, which are fairly portable across areas, comes from our geographic and earnings analysis. The vast majority of variation in skill demand across major-MSA cells is accounted for by major, whereas a much smaller share is accounted for by MSA. Nonetheless, there are substantial remaining cross-area skill differences even within majors. However, cross-area skill differences within majors have only a weak relationship with major earnings premia across areas. Fixed effects for majors explain a considerable share of the variation in cross-cell wages and greatly diminish the predictive power of the individual skill composites. For instance, cognitive, financial, and project management skills are strongly positively associated with cell-level wages, but these patterns are fully accounted for by college majors. This strengthens our conclusion that majors can be thought of as a portable bundle of skills.

Our work contributes to the intersection of several strands of literature. First, we contribute to the broad literature that explores variation in skill demand across firms, markets, and time (Deming and Kahn, 2018; Hershbein and Kahn, 2018). Most work on the supply of college majors focuses on skill-major linkages through occupation (Altonji et al., 2016b; Long et al., 2015). However, occupations are heterogeneous bundles of skills and tasks, and skill demand can vary dramatically across jobs within occupations (Busso et al., 2020). Our analysis highlights the importance of college major as a measurable dimension along which skill demand varies separate from that mediated by occupation.

A second strand of literature looks at whether majors are general versus specialized, which has implications for their returns over the lifecycle. Prior work has examined the benefits of a general versus specialized curriculum in the labor market (Hanushek et al., 2017; Deming and Noray, 2020; Martin, 2022). Several papers do this by quantifying the link between majors and occupations (Altonji et al., 2012; Li et al., 2021; Ransom and Phipps, 2017) or via variation in major premia across occupations (Kinsler and Pavan, 2015; Leighton and Speer, 2020). Our approach abstracts from concerns about selection of college graduates into occupations by using information from job ads prior to employment and realized earnings. Thus, we look at the specific skills associated with each major as perceived by employers and view our approach as complementary to these occupation-based approaches. Our description of the skills employers associate with college majors illustrates one source of the large returns to college major (Arcidiacono, 2004; Kirkeboen et al., 2016; Andrews et al., 2017; Martin, 2022) as well as differences in cost of producing them (Hemelt et al., 2021b).

Finally, we contribute to the understanding of spatial differences in wages, particularly cross-area major wage premia (Ransom, 2021) and spatial differences in the returns to education (Black et al., 2009). In contrast to Deming and Kahn (2018), who find that employer skill demands predict occupational wage premia across areas, we find minimal association between skill demand and cross-area major wage premia. Cognitive and social skills in particular have minimal association with major premia across areas, in contrast to findings for occupational wage premia. This suggests that spatial variation in wages is driven by factors other than within-major skill specialization, at least at the level of aggregate skill composites.

The rest of this article proceeds as follows. Section 2.2 describes the data and sample. Section 2.3 details the relationship between majors and skills. In Section 2.4 we document the geographic variation in the skill-major linkage and then relate skill variation to earnings variation. Section 2.5 concludes.

2.2 Data and Samples

2.2.1 Job Ad Data

We use the near universe of all online job ads posted in the United States from 2010 to 2018, obtained from Burning Glass Technologies (BGT or Burning Glass). BGT scours about 40,000 online job boards and company websites to aggregate job postings, parse and deduplicate them into a systematic, machine-readable form, and create labor market analytic products. The data contain detailed information on over 70 standardized fields including occupation, geography, skill requirements, education and experience demands, and firm identifiers. There are over 15,000 individual skills standardized from the open text in each job posting. Our data cover the United States and contain approximately 153 million individual job postings.

Since the database covers only vacancies posted on the internet, the jobs are representative of a subset of the employment demand in the entire economy. Hershbein and Kahn (2018) conduct a detailed analysis of the industry-occupation mix of vacancies in the BGT data for years 2010-2015 and compare the distribution to other data sources including JOLTS, the Current Population Survey, and the Occupational Employment Statistics. Their analysis suggests that although BGT postings are disproportionately concentrated in occupations and industries that typically require greater skill, the distributions are relatively stable across time, and the aggregate and industry trends in the number of vacancies track other sources reasonably closely.[3] Moreover, since we focus on job ads requiring a bachelor's degree, the skill skew is of even less concern.

[3]See online Appendix A of Hershbein and Kahn (2018).

2.2.2 Sample

We restrict to job postings that list at least one skill, require exactly 16 years of education (i.e., a bachelor's degree), and list at least one college major. Importantly, just over half of the job postings that demand 16 years of education and at least one skill also explicitly list at least one college major.[4] These education and skill requirements leave 12.8% of the original 153 million job postings. Most of our analyses also restrict the sample to ads that list at least one college major posted in metropolitan statistical areas (MSAs). This additional requirement reduces the analytic sample to about 18.5 million unique job postings.[5] We exclude ads specifically targeting workers with graduate education as we are interested in measuring the association between undergraduate majors and skills. In addition, most job postings require 0-5 years of experience, which is more relevant for individuals prior to graduate education.

Given the large reduction in the sample size after imposing these restrictions, one might worry that the types of job postings in our restricted sample differ from the set of all job postings. Table 2.1 compares the occupational composition of job postings in our analytic sample to two larger samples. Differences are mostly due to the bachelor's education requirement. It is well documented that typical job tasks performed in occupations that employ workers with less formal education differ from those that employ workers with more formal education (e.g., Acemoglu and Autor (2011)). The higher concentration of job postings in Management (22% vs. 12%) and Business (15% vs. 7%) occupations in our analytic sample relative to all job postings concurs with this stylized fact. Analogously, the full sample of ads has a higher proportion of job postings in Food Prep (3.4% vs. 0.23%), Building Cleaning and Maintenance (1.11% vs. 0.04%), Sales occupations (11.76% vs. 4.38%), and Office & Administrative Support (9.96% vs. 3.02%).

While the occupational distribution of job postings in the analytic sample (column 5 of Table 2.1) is similar to that of the broader sample requiring 16 years of education and at least one skill (column 3), there are still a few differences of note. The latter sample has a higher proportion of ads listing Education/Training/Library Occupations (2.5% vs. 1.3%), Protective Service occupations (0.3% vs. 0.2%), Sales occupations (8.2% vs. 4.4%), and Office/Admin Support (4.3% vs. 3.0%), with lower proportions in Computer/Math (22.1% vs. 25.8%) and Architecture/Engineering (6.7% vs. 9.3%). This pattern suggests that ads that

[4] Approximately 17% of all postings ask for 12 years of education, 5% ask for 14 years of education, 3% are for 18 years and 1% ask for 21 years of education. The remaining postings are missing information on education (roughly 50% of all postings). For postings that demand 18 years of education, a major is listed as frequently as in postings that demand 16 years of education (54%) but majors are less frequently listed in postings that specify 12, 14, or 21 years of education (6.5%, 37%, and 46%, respectively).

[5] The vast majority of postings are from metropolitan statistical areas, so this restriction drops only about 5% of the "education 16" sample with at least one major (around 1,000,000 postings).

list a college major on average call for occupations associated with higher pay than those that do not.

We more formally investigate these differences using a 1% random sample of job postings that demand a college degree. We regress a binary indicator for whether a job posting lists at least one college major on 900+ metro- and micro- statistical area fixed effects, 99 year-by-month fixed effects, more than 500 six-digit occupation codes, and more than 90 two-digit industry codes. The baseline model, which includes roughly 1,600 covariates, explains only 13% of the variation in whether a job posting lists a major. The explained variation doubles when we include a cubic for the number of skills per posting, indicators for eleven skill composites (described below), and indicators for whether a posting has each of the 1,000 most frequently listed skills. Individually controlling for the 9,000 most frequent skills increases the explained variation to just 29%.[6] These results suggest that differences in extremely detailed observables explain only a modest share of the variation in whether a job ad lists a college major. While our findings rely on the sample of job ads that explicitly list a college major, the degree of unexplained variation in listing a major hints at idiosyncratic reasons for including a major on a job ad. It is thus plausible that our findings would apply to the broader sample of job ads that require 16 years of education. In addition, we assess the robustness of our measures of specificity of skills and majors to the inclusion of ads that do not explicitly list a desired college major.[7]

2.2.3 College Majors

Among job postings that require exactly a bachelor's degree, 54% also list at least one college major. While the exact method used to extract majors from job ads is proprietary to Burning Glass, our discussions with them suggest they do minimal cleaning or imputation beyond standardizing majors into consistent categories. Majors are coded into the Classification of Instructional Programs (CIP) taxonomy at up to six digits (though some ads are initially coded with less granularity), which we first aggregate into four-digit CIP codes. Importantly, a job ad can list multiple college majors. On average, the number of majors listed per ad (conditional on having at least one) remains fairly stable across the analysis period at around 1.7, with about 55% of postings listing a single major, 30% listing two, and 15% listing three or more. For the purposes of analyzing skill demand by major, we

[6] Appendix Table B.2 shows these results. Appendix Table B.3 reports F-tests on the blocks of covariates in the baseline model and reveals that job postings that list a major differ in terms of occupational distribution, industry, and location.

[7] In related work, we are applying machine learning methods to estimate the full latent distribution of majors demanded in job postings.

further aggregate college majors into 70 categories.[8] We aim to produce categories that have meaningful quantities of both job ads (BGT) and degrees granted according to IPEDS. We use the CIP coding hierarchy wherever possible and combine majors that tend to appear in ads together or that require similar sets of skills (as indicated in the job ads).[9] Figure 2.1 plots the share of job postings that list the 10 least and most common majors under this broader method of aggregation. Five majors appear in at least 10% of postings in the analytic sample, including both Business and Computer and Information Sciences, which are listed on 29% and 26% of unique job postings, respectively. The frequency of the remaining 65 majors is quiet heterogeneous, with half of all majors showing up on less than 0.5% of job ads. The least frequently demanded majors in our sample include Theology (0.07%), Atmospheric Sciences and Meteorology (0.03%), Other Physical Sciences (0.03%), and Philosophy and Religion (0.02%).

Since the college majors listed on these job postings have received little scrutiny, an important but open question is how major-specific demand measured in these job postings relates to the composition of bachelor's degrees granted or supplied over time. Figure 2.2 compares the distribution of majors listed on job postings in the BGT data to the distribution of degrees granted for the same majors in the U.S. from years 2010-2018 using IPEDs data. Majors for which the share of job postings is proportional to the share of degrees granted should fall on the 45-degree line, majors overrepresented (underrepresented) in the BGT data will fall above (below) the 45-degree line. Some majors, including Nursing and Economics, have demand that is proportional to the number of degrees awarded for the major. Engineering and Statistics, however, are overrepresented in the BGT data relative to degrees granted, whereas Philosophy and Religion, Atmospheric Sciences, and English are underrepresented.[10] This discrepancy likely reflects a disconnect between the supply and demand for specific college majors, an important topic beyond the scope of this current paper, rather than an issue with the representativeness of the job postings data itself.

[8]There is a 71st category which contains majors that we omit from our analysis. This category contains college majors that are traditionally sub-baccalaureate or remedial programs (e.g., Basic Skills and Developmental/Remedial Education), that are predominantly post-baccalaureate or graduate programs (e.g., Residency Programs), or trade specific (e.g., Mechanic and Repair Technologies/Technicians).

[9]Our process for aggregating college majors is described in Appendix B.1. The full list of all major groups is reported in Appendix Table B.5.

[10]A similar pattern of over- and under-representation is apparent if, instead of IPEDS, we measure supply using the distribution of prime-age workers in the U.S with degrees as measured on the 2009-2018 waves of the ACS.

2.2.4 Categorizing Skills

Burning Glass parses over 15,000 individual skills from the job postings. We categorize by hand the 1,000 most frequent skills into 11 mutually exclusive skill composite categories. To do so, we crafted detailed definitions of the skill composites (see Table 2.2) and then had pairs of our research team manually assign a subset of the skills to one of the composites, using a pre-set process to resolve discrepancies. (We describe the procedure in detail in Appendix B.2.)

This approach provides a few benefits over the application of the keyword approach from Deming and Kahn (2018) or Hershbein and Kahn (2018).[11] First, some of the most frequently listed individual skills are not captured by any skill composite using the keyword approach. Examples include planning (appears on 20% of postings), organizational skills (16%), detail-oriented (12%), scheduling (12%), building effective relationships (11%), creativity (10%), troubleshooting (6%) and multi-tasking (8%). Second, the keyword approach can result in the misclassification of some broad groups of skills. For example, the composite "people management" includes the keyword "management" and thus captures a wide variety of general management activities that do not specifically pertain to managing people, including account management, pain management, and operations management. Similarly, underwriting is also included in the writing composite using the keyword approach, even though that skill is quite distinct.

Table 2.2 provides a description of each of the 11 categories along with the most frequent skills in each category.[12] The final column lists the words used to define these categories based on the keyword approach. Our resulting skill composites are mutually exclusive at the skill level–that is, a detailed skill maps to at most one composite–but a given job posting (or major-by-job posting) can reflect multiple skill composites. Figure 2.3 shows the share of all ads containing a skill falling in each of the 11 categories. "Cognitive" skills are listed in more than three-quarters of all job ads and constitute the most frequently occurring composite (aside from the "unclassified" group, which picks up any skill outside the 1,000 most frequently occurring). In contrast, "people management" and "writing" are the least likely to appear, each mentioned in about one-third of all ads. We note that a much higher share of ads fall into our skill composites than those used by Deming and Kahn (2018), since

[11] In Appendix B.3, we conduct a thorough analysis of the differences between the keyword approach used in Deming and Kahn (2018) and Hershbein and Kahn (2018) and our hand-coding approach. While the keyword approach categorizes more total skills into composites, it misses many relevant and frequent skills, and also results in some inconsistent categorizations. Nonetheless, our results largely hold under either method of constructing skill composites.

[12] Our main analysis focuses on 11 skill composites. In some tables or figures we also provide results for a twelfth skill, communication skills (which is a proper subset of the "social" composite), and a thirteenth composite, unclassified–which consists of all skills outside the 1,000 most frequent.

we have explicitly categorized the 1,000 most frequently occurring skills. Their estimate of the shares of ads seeking cognitive and social skills were 37% and 36%, respectively.[13]

2.2.5 Inferring Desired Skills from Co-Listing with Majors

Our approach assumes that employers list all appropriate skills alongside majors, instead of listing majors in place of desired (or assumed) skills. If employers choose to list a desired major instead of listing the constituent skills, then our metrics will understate the importance of these core skills to a given major. This does not seem to be the case; the most frequent skills appearing alongside majors tend to be core skills required by the jobs these majors tend to enter (Appendix Table B.6). For instance, the top skills for Economics majors include "Microsoft Excel" and "research," those associated with Teacher Education majors include "early childhood" and "child development," and Journalism majors are expected to have "writing" and "editing" skills. Further, when we look at ads for individual occupations, the listed skills tend to be similar regardless of whether a major is listed or not. For example, the top 10 most frequently listed skills on job postings that list the occupation "Managers, All Others" are nearly identical between postings that list a major and those that do not, as are the shares of postings listing each of these skills. This conclusion generally holds for other occupations we examined, including Healthcare and Social Workers, Computer Programmers, Accountants and Auditors, Mechanical Engineers, and Registered Nurses.

Finally, it does not appear that employers are more prone to list a desired major instead of skills in cases where the major has very specific training for particular occupations. While it is true that postings for these majors tend to list fewer skills, there is an extensive amount of variation across majors and even among the more specific majors. For example, postings for Theology majors on average list 6 skills, those for Nursing and Social Work list an average of 10 skills, and those for Electrical Engineering, Business, and Biochemistry & Molecular Biology average 15-17 skills.

Hence, we conclude that employers do not simply list majors as a substitute for listing the skills they seek in job applicants. This pattern is consistent with employers facing a fixed cost of posting a vacancy, but relatively low marginal cost of including additional information like major.[14] The benefits of listing additional information on a posting, even when this additional information is closely related to other material already on the postings (e.g., Teacher Education major and Teaching skill), appear to exceed the costs.

[13] We note that their sample was restricted to professional and managerial occupations but not restricted by education. Our sample is restricted to ads requiring exactly 16 years of education but is not restricted by occupation.

[14] Online postings are likely to be quite different from print job ads in this regard.

While job postings illustrate differences in the *types* of skills associated with each major, we are unable to infer differences in the *level* of skill demanded within each type; wage information attached to the ads is uncommon and likely not representative. Two positions both seeking applicants with "writing" skills may require quite different levels of this skill (e.g., jobs for Journalism majors require more advanced writing skills relative to jobs for other majors). Furthermore, the composite skills we construct also likely mask differences in skill intensity that may be reflected in the detailed set of skills. In either case, to the extent we understate differences in the intensity of skill demand across majors, the large cross-major differences documented below are likely conservative.

A final consideration is that students of varying levels of general ability sort into different majors Paglin and Rufolo (1990); Arcidiacono (2004). Skills stated in job ads may thus reflect employers' perceptions of student sorting, perceptions of human capital accumulation, or both. We do not take a stand on this distinction; either interpretation reflects employers' views of the skills they expect applicants from each major to possess. How intensity of skill level within type of skill can be inferred from job ads is an important direction for further research.

2.2.6 Earnings by Major

To measure average earnings by major across space, we combine the 2009-2018 waves of the American Community Survey (ACS) to create earnings measures at the major-by-MSA level. The baseline sample includes individuals aged 25-54 with at least a bachelor's degree. We drop observations with imputed or negative earnings or imputed majors. We keep all individuals with positive years of potential experience and positive weeks worked. Finally, we impose the additional restrictions that workers are not enrolled in school and are full-time, full-year workers (FTFY), where full year is defined as at least 40 weeks a year and full-time is defined as 30 hours a week.

We construct hourly earnings by dividing annual earnings by the product of weeks worked during the past 12 months and usual hours worked per week. We adjust earnings for inflation to 2019 dollars using the Personal Consumption Expenditures (PCE) deflator from the Bureau of Economic Analysis (BEA). In our analyses, we use two versions of real hourly earnings. The first is the log of raw mean hourly earnings in the major-MSA cell. For the second, we regression-adjust for compositional differences across majors. Specifically, we regress the log of hourly earnings at the individual level on indicators for female, Black, and Hispanic, as well as a quartic in potential experience, and we then take the mean of the residuals within each major-MSA cell.[15] Figure 2.4 shows substantial geographic variation both across and within

[15]In both cases we employ sample weights when aggregating to major-MSA cells.

majors in the mean hourly wage of full-time, full-year, prime-aged workers in the United States. We later assess the extent to which this variation can be explained by differences in the skill content across and within majors.

2.3 Skills Associated with College Majors

Table 2.3 reports the share of ads listing each of the skill clusters separately for a handful of majors, along with the minimum and maximum share across 70 different majors.[16] There is a substantial range across fields for many of these skill aggregates. For instance, the share of ads desiring specific software skills ranges from less than 4% for Nursing to (unsurprisingly) nearly all job ads in Computer Science. Project management skills are sought in nearly all job ads for Public Health majors but rarely for jobs seeking Education or Foreign Language majors. People management is rarely desired on job ads associated with Accounting majors, but appears on more than half of ads targeting Public Administration majors. Because "communication skills" constitute such a large share of the "social skills" composite, we separately report statistics for this skill.

2.3.1 Measuring Skill Content

We formalize this variation in skill demand across majors in two ways. First, we construct a Location Quotient (LQ) for each major-skill-composite combination. This measure is commonly used to characterize the concentration of industry- or occupation-specific employment in a region relative to the nation. The LQ is the ratio of the demand for a skill among job postings listing a particular major relative to the demand for that skill among all job postings. For the dyad of major m and skill component s, the LQ is computed as:

$$LQ_{sm} = \frac{(N_{sm}/N_m)}{(N_s/N)} = \frac{(N_{sm}/N_s)}{(N_m/N)} \tag{2.3.1}$$

where N_m is the number of ads that list major m, N_{sm} is the number of ads that list major m and skill s, N_s is the number of ads that list skill s, and N is the total number of ads. In our main specification, we measure national skill demand (also referred to as the market demand) using all postings that require 16 years of education and list at least one college major. We construct one LQ for each skill composite s and major m combination. An LQ around 1 indicates that the demand for a skill among job postings with major m is the same as the market demand for that same skill. An LQ>1 indicates that the skill is concentrated

[16] Full results for all 70 majors are in Appendix Table B.7.

among ads that list major m because the fraction of ads demanding the skill in the entire market is lower than the fraction of major m ads listing that skill.

One complication in practice is that a job posting can list multiple majors and multiple skills; this is not an issue in more commonly used settings in which the allocations of workers to occupations and regions are mutually exclusive. In the common setting, regional employment sums to national employment, and the occupation-specific employment in a region sums to total regional employment. As a result, the average of occupation-by-region LQs for a given region weighted by the occupation's share of national employment for each region equals one. In our case, because we treat a single job posting that lists X different majors as X different observations, the above properties no longer hold, muddying interpretation of the LQ.

To recover the desirable properties of LQs, we make a few adjustments. First, we redefine the total count of job postings (N) to be the total number of job-posting-by-major observations ($\hat{N}$) so that $\sum_m N_m = \hat{N}$. Second, we analogously redefine the total count of unique job postings with skill s (N_s) to be the total of job-posting-by-major observations that list skill s ($\hat{N}_s$) so that $\hat{N}_s = \sum_m N_{sm}$. With these changes, the adjusted LQ for a dyad of major m and skill s is:

$$L\hat{Q}_{sm} = \frac{(N_{sm}/N_m)}{(\hat{N}_s/N)} = \frac{(N_{sm}/\hat{N}_s)}{(N_m/\hat{N})} \tag{2.3.2}$$

The distribution of the adjusted LQs across majors for a given skill now has a weighted average of 1, where the weights are equal to the shares of all job-posting-by-major combinations that list major m. As a result, we can compare the adjusted LQs to 1 to determine relative concentration.

To characterize the degree of specialization of a major as reflected by the skill composites, we examine whether a major has LQs close to 1 for each of its skill composites. Specifically, for each major, we compute the absolute value of the deviation of each skill composite LQ from 1. We then sum the absolute value of the deviations within major and across all 11 skill composites: $\sum_{s=1}^{11} abs(\widehat{LQ_{sm}} - 1)$. Majors with a higher sum are more specialized.

Our second approach compares the skills demanded from each major to national skill demand using a cosine similarity measure and the 9,000 most frequently listed skills.[17] Specifically, for all job ads in the national analytic sample and for ads listing each of 70 different majors, we construct a vector containing the share of all ads listing each of the 9,000 skills. We then construct the cosine similarity between the national skill distribution and major-specific distributions. We measure the distance between a major's 9,000-dimensional

[17] We narrow our focus from the complete set of 15,000 skills to the roughly 9,000 skills found on at least 0.001% of all job postings.

skill demand vector and the 9,000-dimensional national skill demand vector using the angle between the two vectors. Majors with a value closer to zero have skill demand that is very different from national demand and are thus more specialized, whereas more general majors with a skill demand vector that is similar to the national vector will have a cosine similarity near one.

The cosine similarity and LQ measures of skill concentration provide complementary information. The former measures how similar a given major is to the broad set of jobs based on nearly the entire skill vector, which includes many infrequent and specific skills. In contrast, the latter focuses on similarity based on the large clusters of the most common skills. The LQ-based measure also permits us to characterize skill differences across majors along a tractable number of dimensions. We assess the empirical correspondence between these two measures in a subsequent section.

2.3.2 Skill Specificity of College Majors Based on Location Quotient

Across the 70 majors and 11 skill composites, we construct nearly 800 different LQs, one for each skill-by-major combination. The first row of Table 2.3 reports the denominator of the LQ for each skill composite, which is roughly equivalent to the percentage of job postings that list each skill. In Table 2.3, for a selected set of majors, we list the share of each major's postings that list each skill. This term is the numerator of the LQ, and is particular to a given major-by-skill combination. The LQ is simply the ratio between each subsequent row and the top row.

We summarize our findings from the LQ calculations graphically. Panel A of Figure 2.5 plots the distribution of LQs across majors for four skill composites. Social and organizational skills have a large number of major-specific LQs that are clustered around 1, indicating that most majors require similar levels of these skills. Customer service and financial skills are more varied; some majors are associated with very high levels of those skills (such as Social Work and Construction Management, respectively) and others very low (Atmospheric Science and Theology). Panel B combines the LQs into a single index–the share of the LQs that are within narrow bounds around 1–which measures the specificity of skills to majors. For a given skill, if most majors have an LQ around 1, then the demand for that skill is not particularly concentrated among any subset of majors. Most majors have an LQ for social skills near 1 because most majors have the same fraction of ads demanding social skills as does the entire market. Social skills are thus general–a skill that is demanded across ads for most majors. In contrast, Financial and Customer Service skills are specific.

Figure 2.6 plots the LQs for all majors and the 11 skill composites. Majors are ordered according to the degree of overlap between a major's skill demand and national demand. For each skill composite, we measure the absolute deviation of the major's LQ from one, and then sum the absolute deviations across all skills for a major.[18] For some majors, including Business, Economics, and General Engineering, the measure is very small, suggesting that they have a skill profile similar to that of the broader job market: LQs fall close to one for all skill aggregates. These majors can be thought of as *general* in the sense that they are associated with skills that are demanded by a large number and wide variety of jobs in the college-educated labor market.

Majors towards the bottom are specialized in the sense that they reflect a skill profile that is quite distinct from the labor market overall. These include Nursing, with a high co-occurrence with customer service but very low with software, computers, financial, and writing. Among postings that demand a Nursing major, 23% demand computer skills, which is roughly half the market-wide demand of 42%, yielding an LQ of 0.5. The demand for writing and software skills for Nursing is even lower. A desire for customer service skills, however, is overrepresented: they appear on 82% of postings that list a Nursing major but only 46% of job postings in the wider sample. Foreign Language has a high concentration of social skills and writing but low need for software or financial skills.

Majors in the middle, such as Computer Science and Psychology, have a skill profile broadly reflective of the national one, but with a few skill categories that are particularly over- or underrepresented.

These results are robust to including postings that demand 16 years of education but do not list a major when calculating the LQ denominator. Our main measure compares the share of each major's postings that list each skill to the percentage of all job postings with a college major that list each skill. However, it is possible that the postings that do not explicitly list a college major are searching for workers with any disciplinary training. If so, then the skill demand on these postings represents the skills employers expect the average college graduate to possess. To assess this, we reconstruct the LQ measures with all postings that demand exactly 16 years of education (irrespective of whether a major is listed) in the denominator. The ranking of college majors is almost identical to our preferred specification ($R^2 > 0.95$).

2.3.3 Measuring Specificity with All Skills

We also compare our LQ-based measure to the cosine similarity measure. The cosine similarity metric captures the similarities between each major and all job ads nationally along

[18] Specifically, for each major, the measure is $\sum_{s=1}^{11}[abs(LQ_{s,m}-1)]$ where the sum is taken across skill composites within a major. We also order majors using the sum of squared deviations $\sum_{s=1}^{11}[(LQ_{s.m}-1)^2]$. The ranking of majors based on the two measures is highly correlated (0.96).

the vector of 9,000 skills, which incorporates more information about less frequent, possibly more specialized, skills. Figure 2.7 shows that the two metrics produce broadly similar rankings of specificity across majors. The R^2 from the bivariate regression between major rankings of the two indices is 0.37 when majors are equally weighted and 0.53 when majors are weighted by the number of ads; the association is similar if we use the metric itself, rather than the rank, as the outcome (Appendix Table B.8). This strong correspondence reflects the fact that most of the variation in the cosine similarity measure comes from variation in the 1,000 most frequent skills ($R^2 = 0.90$), which are the ones that enter our LQ-based index.[19]

Figure 2.8 plots the similarity of skill demand between each pair of majors along the vector of 9,000 skills. Majors that have similar skill demand have a value closer to 1 and are substitutes in terms of skill demand; these are represented by a darker shade. Unsurprisingly, some of the closest major pairs occur within the same broad CIP category, including the pairs of Finance and Accounting; Communication & Media Studies and PR & Advertising; and Statistics and Mathematics. However, close majors are also found across different broad categories of study, including the pairs Other Engineering and Business; and Political Sci/Gov & Int'l Relations and English, Liberal Arts, & Humanities. Finally, some majors have many substitutes, which we proxy by the share of other majors to which the given major is very similar (similarity measure >.8), including Business, Library Science, English, Liberal Arts, Humanities, and Communication & Media Studies.

The graph also clearly highlights specific majors: Teacher Education and Nursing are both represented by light boxes across the graph, as their skill vector is quite different from almost all other majors and they have few substitutes. Both our LQ-based and cosine-similarity-based metrics distinguish general from specific majors, though they use employers' stated skills in different ways. Furthermore, the extent of skill substitutability clearly differs across majors, often in ways not captured by the CIP code classification hierarchy.

2.3.4 Comparison to Prior Work on College Major Specificity

Our measure of college major specificity complements those constructed by other scholars, which primarily rely on major-occupational linkages and earnings premia across majors. Figure 2.9 compares our measure to one based on the occupational concentration of college majors, specifically the share of recent college graduates with a given major represented in the top five most frequent occupations in the ACS. There is a strong correlation between major rankings when cells are weighted by the number of ads (.47), but minimal correlation

[19] In addition, the R^2 from the bivariate regression between major rankings using the LQ-based measure and the cosine similarity measure based on only the 1,000 most frequent skills is almost identical to that yielded when the cosine similarity measure is instead based on the top 9,000 skills.

when they are unweighted (.004), suggesting that inferences about specialization are more robust for more common majors.[20]

Leighton and Speer (2020) construct a Gini coefficient of wage premia across occupations. The notion is that majors with highly occupation-dependent wage premia are likely providing more specialized skills. Kinsler and Pavan (2015) develop a similar idea by focusing on wage differences between workers in jobs that are or are not related to their major. Relatedly, Li et al. (2021) build a complexity measure of majors based on the breadth of occupations to which a major maps and the narrowness of majors that in turn feed into those occupations. Ransom and Phipps (2017) use major-to-occupational flows to construct measures of major occupational "distinctiveness" and "variety." Appendix Table B.9 compares the most/least specific majors using our two skill-based metrics to those published by Leighton and Speer (2020). A few majors appear on multiple lists, most notably Nursing and Education (most specific) and Mathematics (most general).

Thus, there is a correspondence between which majors are considered general or specific when skills are measured based on employers' perceptions as expressed on job postings and when measured based on realized occupational sorting. Our measure of specificity, which is based on skill demand, additionally permits investigation of specific mechanisms that likely contribute to major wage premia–particularly related to the role of geography.

2.4 Skill Variation Across Areas and Earnings Variability

The prior analysis demonstrated the substantial variation in skills associated with college majors, aggregated across all years and labor markets. However, the universality and granularity of the BGT data also enable us to analyze major-specific variation across space; geographic skill variation has been shown to be important for occupations (Deming and Kahn, 2018). In this section, we quantify the extent of variability in skills associated with each major across areas and use this variability to examine how skills and majors relate to earnings. Substantial variation across space in skill demand for the same major may indicate that local postsecondary providers will need to tailor program curricula to suit local labor market needs.

2.4.1 Geographic Variation in Skill Demand

Figure 2.10 depicts variation across the more than 900 U.S. micropolitan and metropolitan statistical areas in the share of job postings for Business majors that seek cognitive skills. Areas with darker shading have larger shares of Business major ads that demand cognitive

[20] Appendix Table B.8 presents correlations between all of the specificity measures we construct.

skills. Contrast Jasper, Indiana and London, Kentucky. Both locations have similar quantities of job postings for Business majors (∼500-700 job postings). However, in Jasper, roughly 82% of job postings for Business majors demand cognitive skills compared to only 46% in London, KY. Even though these two localities are only a 3-4 hour drive apart, employers in these areas demand very different skills from Business majors. Next, beam down to Roswell, NM and nearby Andrews, TX. These locales differ in both the quantity of job postings that list Business majors and the percentage of those job postings that demand cognitive skills.

Table B.2 quantifies the amount of variation in skill demand captured by majors and places. We construct major-MSA cells containing the share of ads seeking each skill. Majors account for the vast majority of the variation across these cells–major accounts for almost 90% of the cross-cell variation in demand for software skills and three-quarters of the variation for customer service skills. Place accounts for only 3-11% of the cross-cell variation in skill demand. The remaining, unexplained variation in major-specific skill demand across areas is substantial–up to 50% for organizational and communication skills.

2.4.2 Skill Demand and Earnings

Is this variation consequential in terms of wages? Figure 2.4 showed substantial wage variation across majors and areas. We now examine whether such differentials are associated with differences in skill demand. Returning to the previous examples, in Jasper, IN, the average adjusted hourly earnings among Business majors is \$44.30, which is about 5% higher than the adjusted hourly earnings of \$41.90 in London, KY, a place where employers demand relatively less cognitive skill of Business majors. The average adjusted hourly earnings in Andrews, TX (\$43.70) are 7.5% higher than in Roswell, NM, also consistent with the relatively higher demand for cognitive skills.

To systematically examine whether skill requirements on job postings are related to earnings, we estimate variations of the following regression model:

$$Y_{jk} = \sum_{s=1}^{S} \beta_s \text{PctSkill}_{sjk} + \gamma_k + \gamma_j + \epsilon_{jk} \tag{2.4.1}$$

where Y_{jk} is the log of mean hourly earnings (2019 dollars) among college graduates in major k in MSA j from the ACS, and PctSkill_{sjk} is a vector of skill demand in the major-MSA cell measured by the share of ads that list each skill. The coefficient β_s indicates the approximate hourly earnings change associated with a 100-percentage-point increase in the share of job ads requiring the skill. The inclusion of major (γ_k) or MSA (γ_j) fixed effects isolates the association between skills and earnings that occurs within majors and MSAs, respectively.

We weight each observation by the number of employed people in each cell using person weights from the ACS.[21]

We report results from our preferred specification in Panel A of Table 2.5. The first model, in column 1, includes only the 11 skill composites and reports the raw correlation between skill demand and log mean hourly earnings in a major-MSA cell. Skill demand is highly correlated with earnings. Major-MSA cells with high demand for cognitive, financial, and project management skills have much higher hourly earnings than those with low demand for such skills. A 10-percentage-point increase in the share of ads demanding cognitive skills is associated with a 4% increase in average wages. Greater demand for people management, social, and basic computer skills (conditional on other skills) are negatively correlated with earnings. These traits may be markers for lower-paid occupations. Collectively the 11 skill composites explain 34% of the wage variation across MSA-major cells and are jointly statistically significant at a 1% level (F-statistic = 17.9, p = 0.000).

Specification (2) includes MSA fixed effects, accounting for any systematic pay or cost-of-living differences that correlate with the skill content of jobs across areas. If in certain MSAs employees are more likely to work in teams, employers will demand more social skills from all majors in the MSA. Alternatively, firms may list more skill requirements in cities that have more skilled workers (Deming and Kahn, 2018). The inclusion of MSA fixed effects accounts for these MSA-level aspects of skill demand as well as pay differences that are due to MSA-wide factors including cost of living. The inclusion of MSA fixed effects does not alter the overall patterns seen in the raw differences. Cognitive, financial, and project management skills are still associated with higher wages. While geographic variation in wages is important–underscored by the near doubling of the explained variation–it is mostly uncorrelated with skill demand among our sample of workers with bachelor's degrees.

Finally, specification (3) adds major fixed effects, absorbing any systematic pay differences across majors that occur in all labor markets. Fixed effects for majors explain a considerable share of the variation in cross-cell wages and greatly diminish the predictive power of the individual skill composites. This suggests that majors can be thought of as portable bundles of skill composites. Once we account for major and MSA, the remaining variation in skill demand measured by the skill composites explains relatively little additional wage variation (F-statistic = 2.8, p = 0.004). As Appendix Table B.2 showed, this is not because there is no remaining variation in skill demand within majors across areas; one-third of the variation in demand for cognitive skills remains in this final regression, but its level does not systematically

[21] Although we mostly focus on weighted regressions, we also estimate models in which each major-MSA combination is equally weighted. Unweighted estimates are generally consistent with weighted estimates, with a few exceptions that we discuss below.

correlate with earnings. The only remaining statistically significant skill-wage correlation is that demand for basic computer skills is associated with lower wages. This association is small in magnitude: a 10 percentage point increase in the share of ads desiring basic computer skills is associated with a 0.5% decrease in average wage.

Panel B of Table 2.5 demonstrates the robustness of these results. We report only specifications that include MSA fixed effects, analogous to specifications (2) and (3) in Panel A. Specifications (4) and (5) adjust wages for individual-level demographics (age, sex, race) before aggregating up to the major-MSA cell level. Specifications (6) and (7) weight each cell equally. Specifications (8) and (9) compute cell-level wages for workers under the age of 35 to better reflect the wages of recent college graduates. The final two specifications, (10) and (11), restrict analysis to job ads that have no more than minimal work experience required in order to reflect entry-level skill demand among college graduates. Across all specifications, results are similar and the qualitative picture does not change. This suggests that the skill-wage relationship we document is not driven by demographics, density of majors, age profiles, or demand for experience by major.[22] The broad patterns hold: skill demand can explain an appreciable share of the cross-cell wage variation, but most of this can be accounted for by major-specific effects. Cross-area variation in composite skill demand within majors, as documented in Figure 2.10, does not correlate with earnings. A caveat, however, is that this analysis is silent about whether variation in *individual* skills within majors across places–as opposed to skill composites– relates to earnings.

This finding stands in contrast to Deming and Kahn (2018), who find that local employer (composite) skill demand predicts wages across areas, even after controlling for occupation and other confounders.[23] In particular, we find that both social and cognitive skills have minimal association with major earnings premia, while Deming and Kahn (2018) find that these skills are associated with area-specific occupational wage premia. Their result suggests caution in interpreting occupations as uniform bundles of tasks: there remains ample variation in skill demand across place and within occupation that is relevant to wages. In contrast, a worker's college major can more reasonably be considered a portable bundle of skills. Differences in skill demand within majors may happen at a much more granular level than the level of aggregation captured by our skill composites. Further, these patterns could also indicate differential sorting of majors into occupations across places. For instance, technology jobs may

[22]Using a wider experience window (0 to 4 years, 0 to 6, etc.) produces very similar results. The vast majority of job ads list minimal experience. Nearly 80% require 5 years or fewer (including 25% that do not require any experience), and only 2% of ads seek more than 10 years of experience.

[23]We attempt to replicate Deming and Kahn (2018) in Appendix B.4. Differences can be explained by some combination of skill classification method (keyword vs. hand-coding the top 1,000 skills), weighting, and manner of aggregation (occupation-MSA vs. major-MSA), with little role for sample differences. Further, we conclude that associations between wages and social skills are especially sensitive to these decisions.

be disproportionately filled by Computer Science majors in Silicon Valley but by Business majors in Scranton.

2.5 Conclusion

In this paper, we provide a comprehensive account of the skills associated with college majors as perceived by employers and expressed in job ads. The choice of field of study during college is one of the most direct ways college-educated individuals acquire skills and signal capabilities to employers. Thus, a more thorough understanding of the relationship that conjoins majors, skills, and jobs stands to inform policy leaders in higher education and industry.

We use data from the near universe of online job postings over the period 2010-2018 to develop measures of skill and major specificity inspired by the logic of location quotients (LQs) from the literature on industry concentration, as well as measures based on cosine similarity to capture high-dimensional vectors of skills. These measures of skill and major specificity complement and extend recent developments in this space (Leighton and Speer, 2020; Li et al., 2021) by focusing on specific skill demand manifested in job ads, thereby allowing us to compute such measures based on information that precedes the employment choices of individuals, a more proximate and direct signal of skill demand independent of occupational sorting.

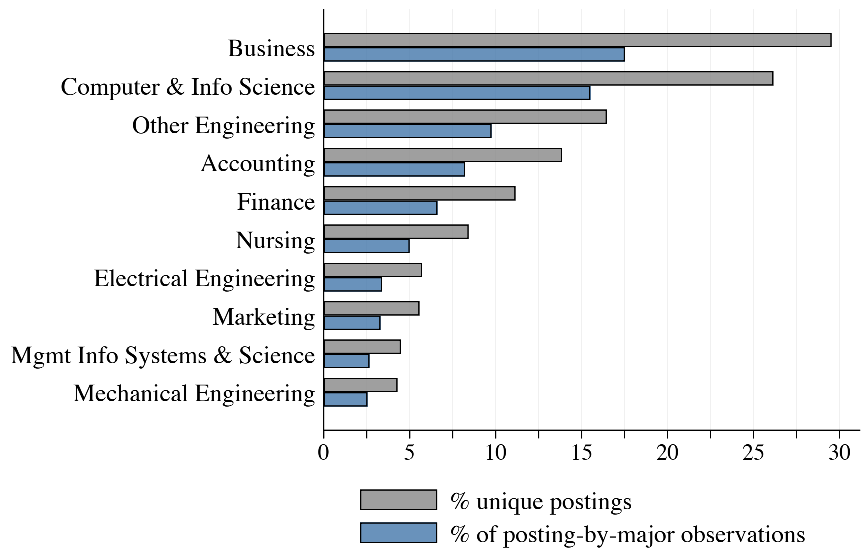
We find that some majors such as Business and Engineering are general due to the fact that demand for most of their component skills is neither under- nor over-concentrated among job ads listing those majors. Other majors, such as Nursing, are more specific in being closely associated with skills that are not widely sought in the labor market for college graduates.

Mapping similarities among majors based on our skill demand measures highlights the fact that common classification systems based on curricula (such as CIP) may not reflect salient dimensions of different fields of study. That is, a student can develop project management skills through interactions with a variety of substantive material–and majors that develop such skills well are likely to have similar labor market payoffs. Hence, one implication is that policymakers and higher education leaders may want to adopt a broader and more multi-dimensional view of how college majors relate to competencies demanded by the labor markets most relevant for their institutions' graduates.

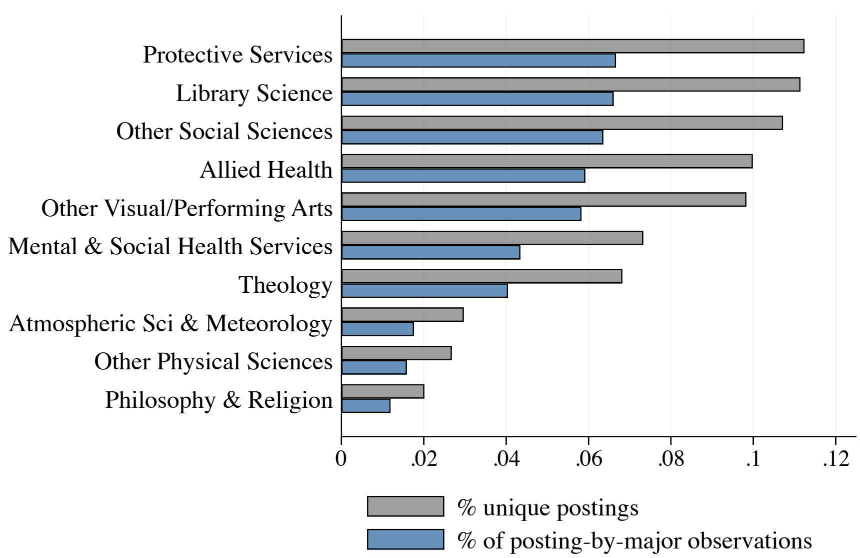
We use information on earnings by major from the ACS to characterize associations between majors, skill demand, and earnings across locations. We document substantial variation across space in both skill demand and average earnings by major. Despite the fact that variation in skill demand remains after accounting for major and geographic location, we find little evidence that such remaining variation meaningfully correlates with variation in earnings. This suggests that majors can generally be conceptualized as bundles of aggregate skills that are fairly portable across areas in ways that occupations are not. However, our analysis leaves open the possibility that a more fine-grained categorization of skills–such as the thousands that are available in job postings–could still matter for explaining wage variation within major and across place. Further analysis of the detailed dimensions of skill demand by college major would add to our understanding of worker-employer matching in the growing labor market for college graduates, and it could also provide better pathways for institutions of higher education to differentiate the skill sets with which they equip particular majors. For example, efforts to adjust the supply of workers with particular skills to meet local employment needs should consider that the hiring decisions of firms depend on their perception of the skills possessed by particular types of workers.

Figure 2.1: Most and Least Frequently Demanded Majors

(a) Most Frequently Listed

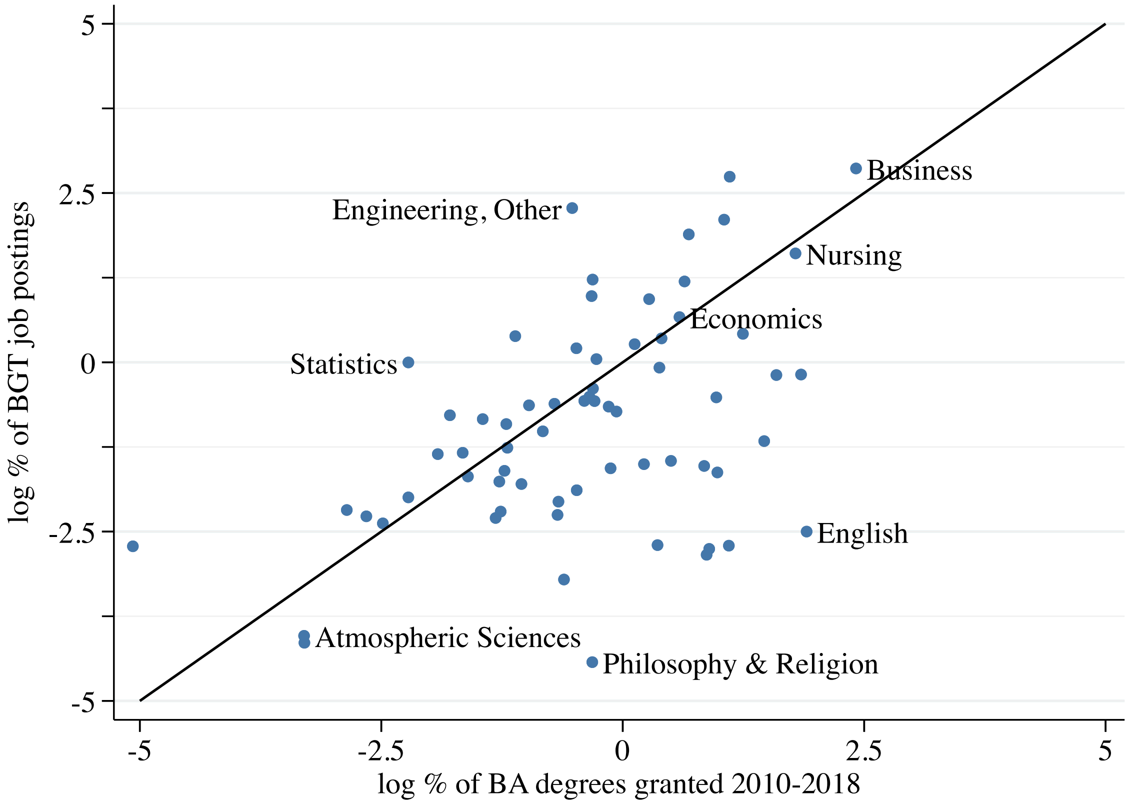


(b) Least Frequently Listed



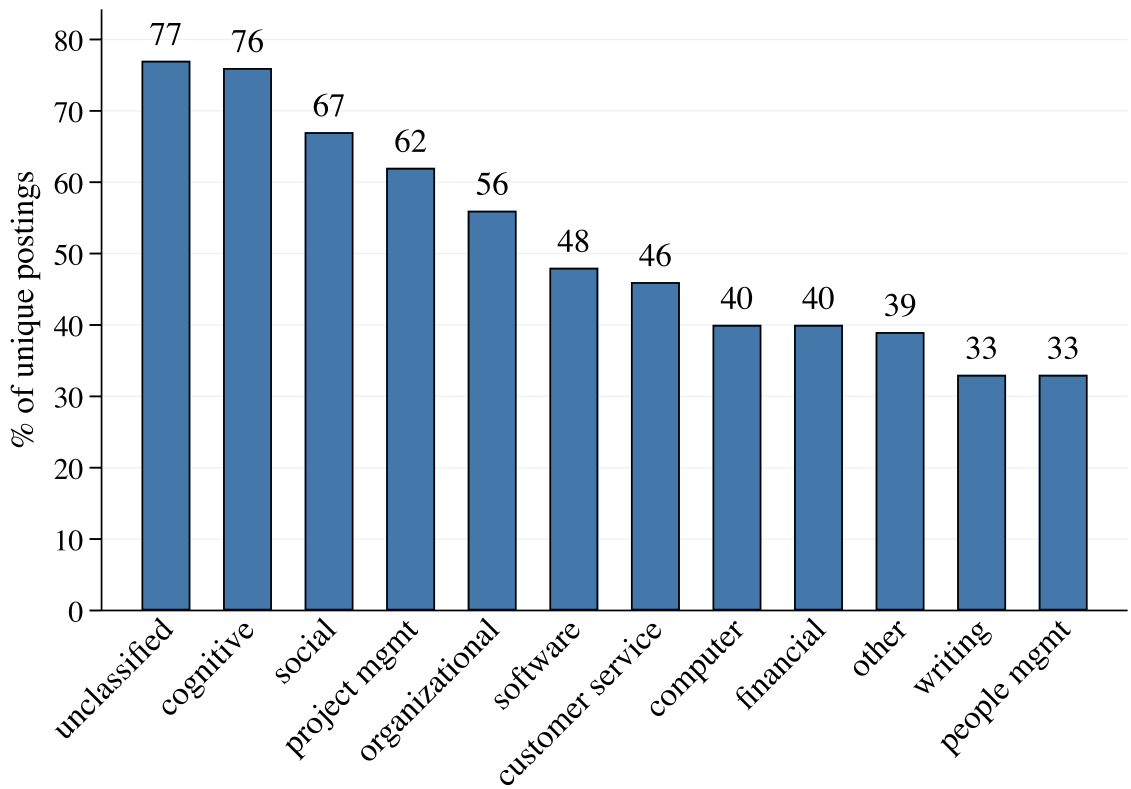
Note: Figure plots the percentage of Burning glass job postings listing each major. Sample includes all job ads posted between January 2010 and May 2018 in metropolitan statistical areas that list 16 years of required education, at least one skill, and at least one major.

Figure 2.2: Comparison between Major Share in Ads vs. BA Completions



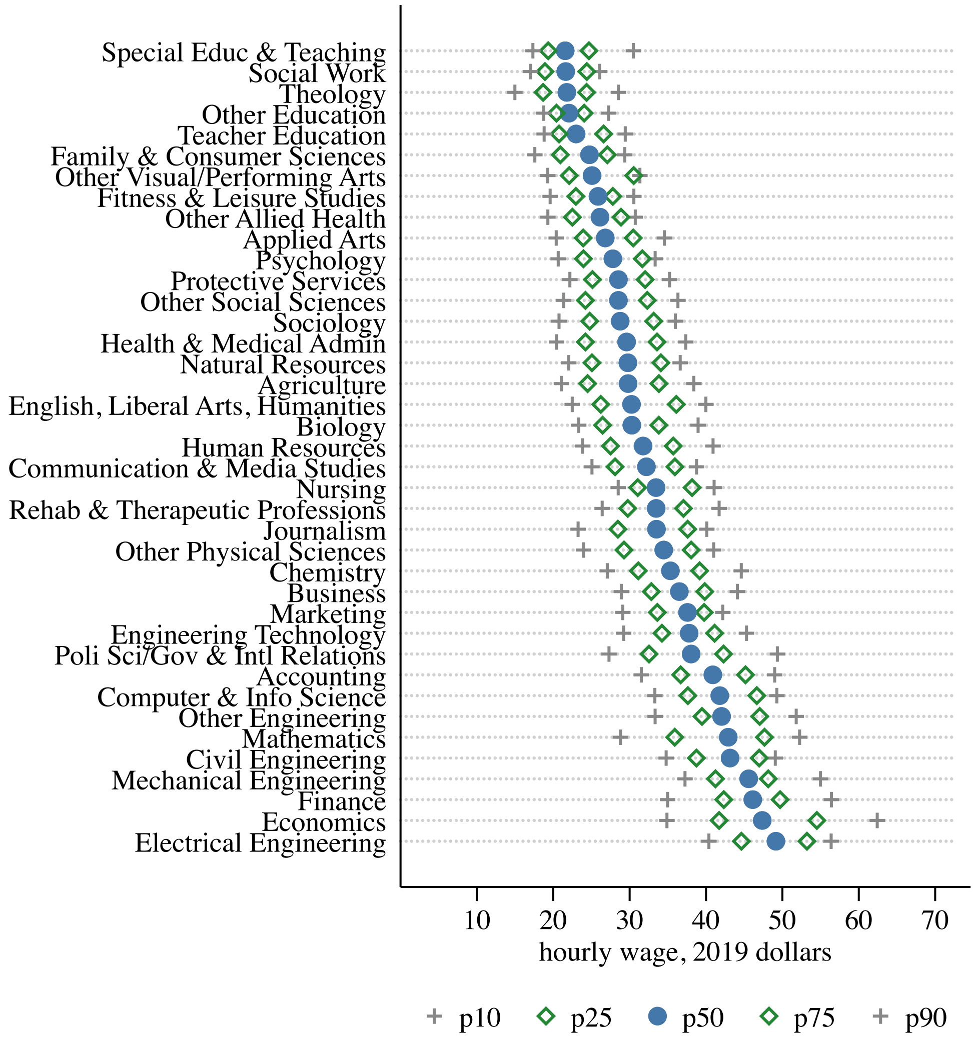
Note: Figure plots the log percentage of Burning glass job postings listing each major against the log percentage of degrees granted (from IPEDs data) in years 2010-2018.

Figure 2.3: Skill Composites: Percentage of Unique Job Postings Containing Skill Composite



Note: Figure plots the percentage of Buring Glass job postings listing at least one skill in each of 11 skill composites constructed from the 1000 most frequent skills. See Table 2.2 for definition of skill composites. A twelfth composite, "unclassified", is the share of ads containing a skill outside the 1000 most frequent. Only 0.2% of postings list none of our 11 composites (excluding "unclassified"). Across job postings, the mean and median number of composite skills listed is five (excluding "unclassified").

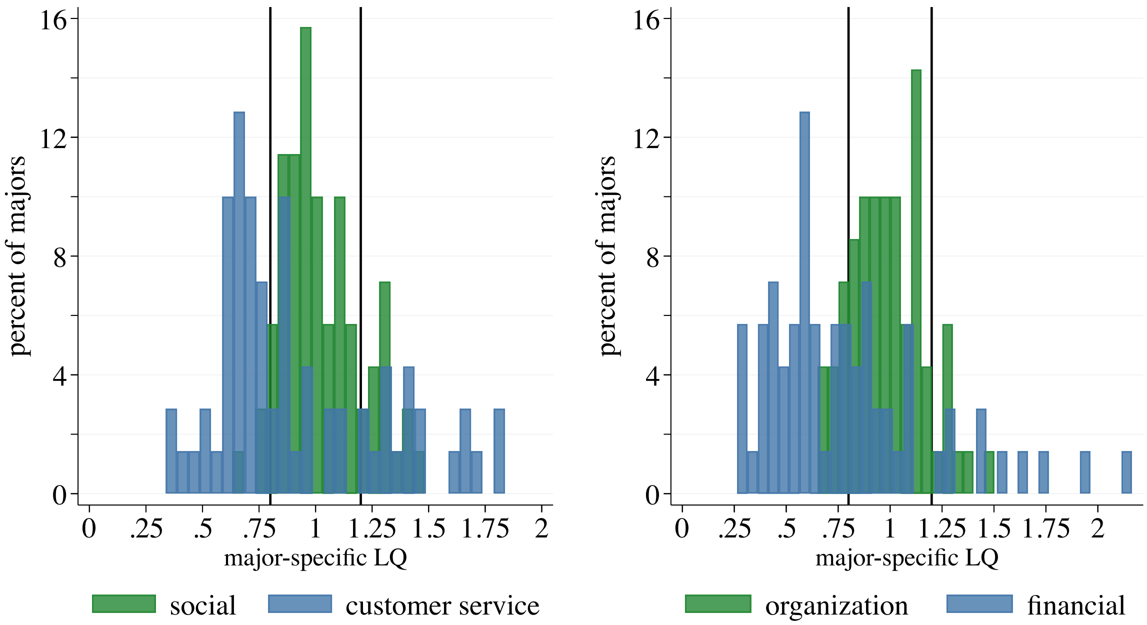
Figure 2.4: Distribution of Average Wage Across Majors and Areas



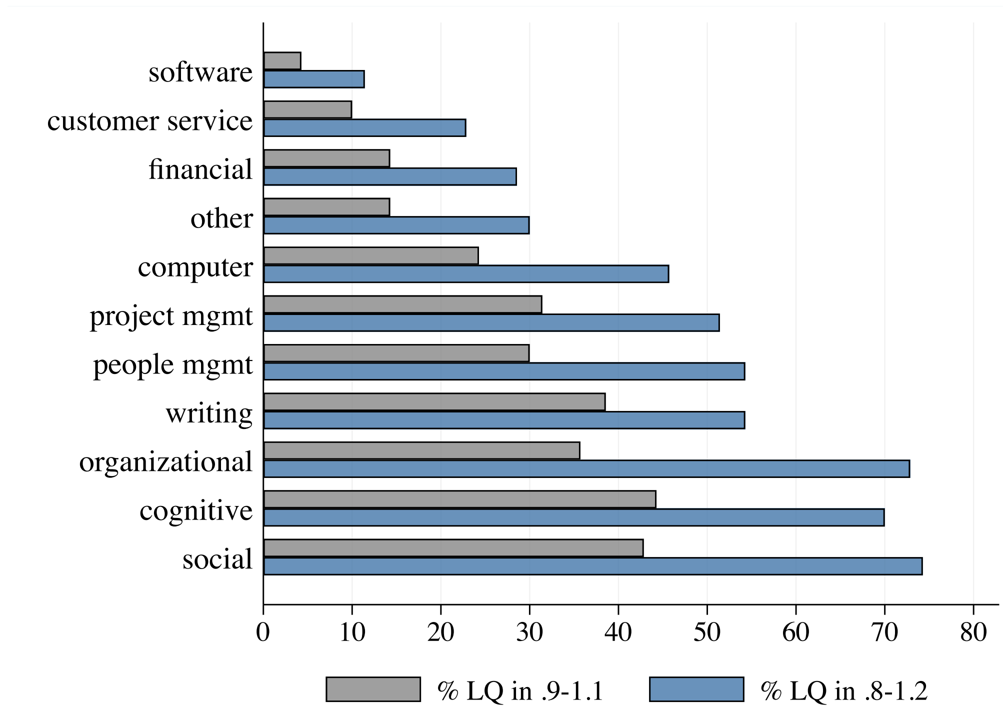
Note: Mean hourly wages for each major-MSA cell in the U.S. are computed from the American Community Survey 2009-2018. Sample includes only full-time, full-year, prime-age workers with exactly a bachelor's degree. Figure includes the 39 majors (out of 70 we classify) with estimates in at least 600 CBSAs (metropolitan and micropolitan areas).

Figure 2.5: Distribution of of Skill Concentration Across Majors

(a) Full Distribution for Four Skill Composites

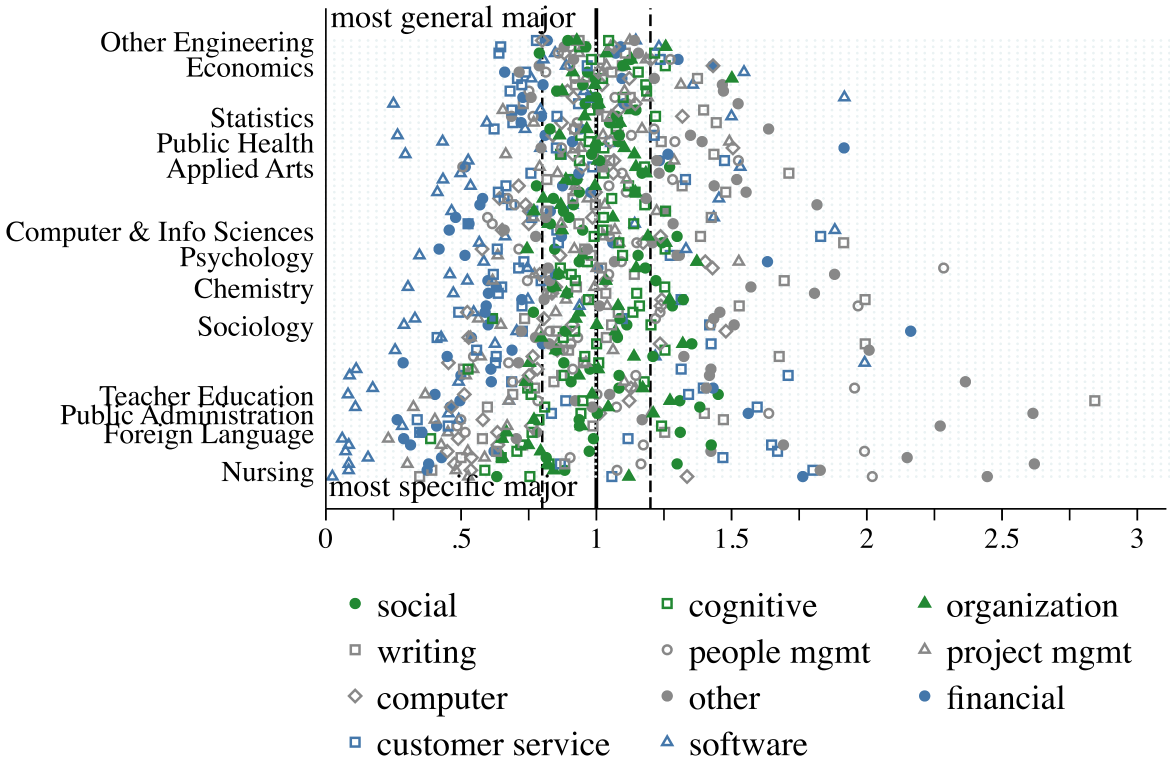


(b) Skills Ranked by Specificity to Major



Note: Panel A plots the distribution of location quotients (LQ) across all 70 unique majors for each of four skill composites as measured in the Burning Glass data. A LQ greater than 1 indicates that ads with a given major are more likely to seek the skill than ads overall. Sample includes 37.1 million major-ad combinations. Panel B plots the (unweighted) share of LQs that are within a narrow range of 1. Lower values indicate skills that are more major-specific.

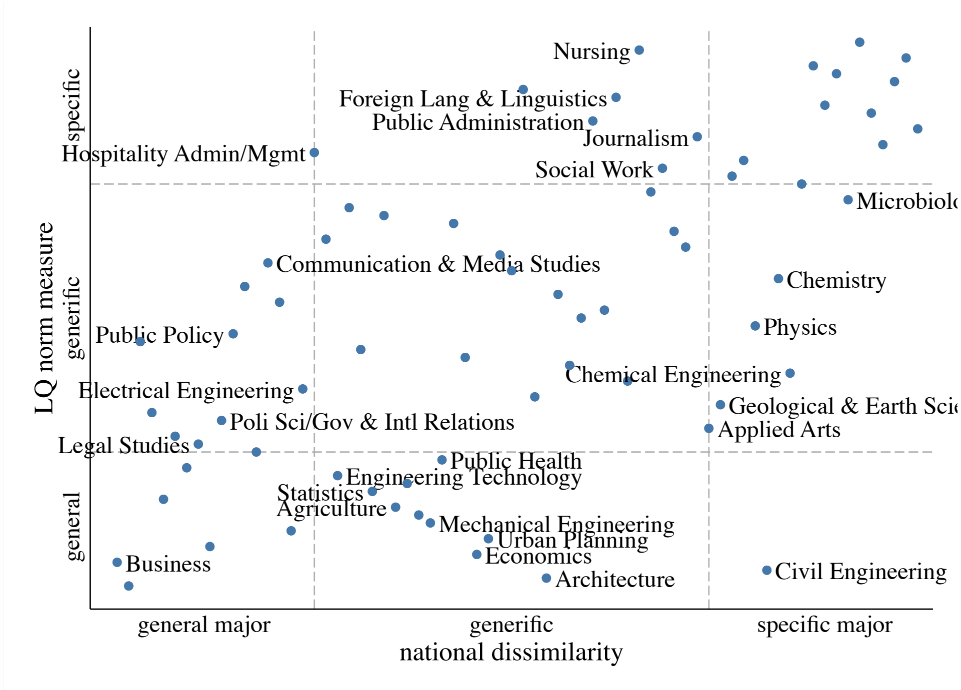
Figure 2.6: Skill Concentration for All Majors



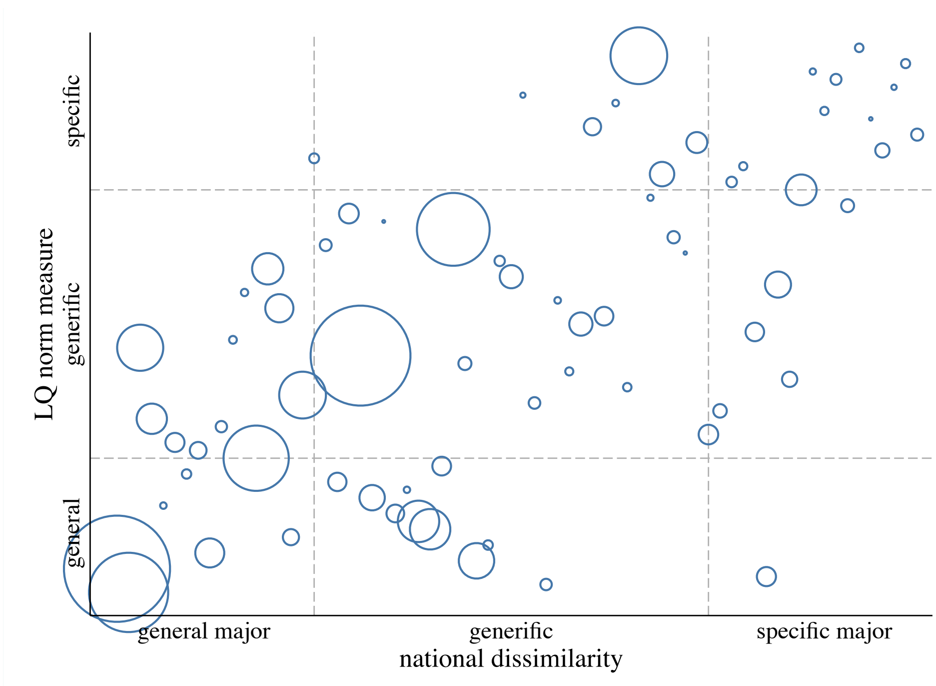
Note: Figure plots the location quotients (LQ) for 11 skill clusters for 70 majors based on skill demand as measured in the Burning Glass data. See text for details on the LQ measure. An LQ greater than 1 indicates that ads with a given major are more likely to seek the skill than ads overall. An LQ less than 1 indicates that ads with the major are less likely to seek the skill than ads overall. Skill composites indicated by green markers are considered more general skills, skill composites indicated by blue markers are specific and skills indicated by gray markers are generific.

Figure 2.7: Skill Composite vs. Similarity Index Measure of Concentration

(a) Unweighted

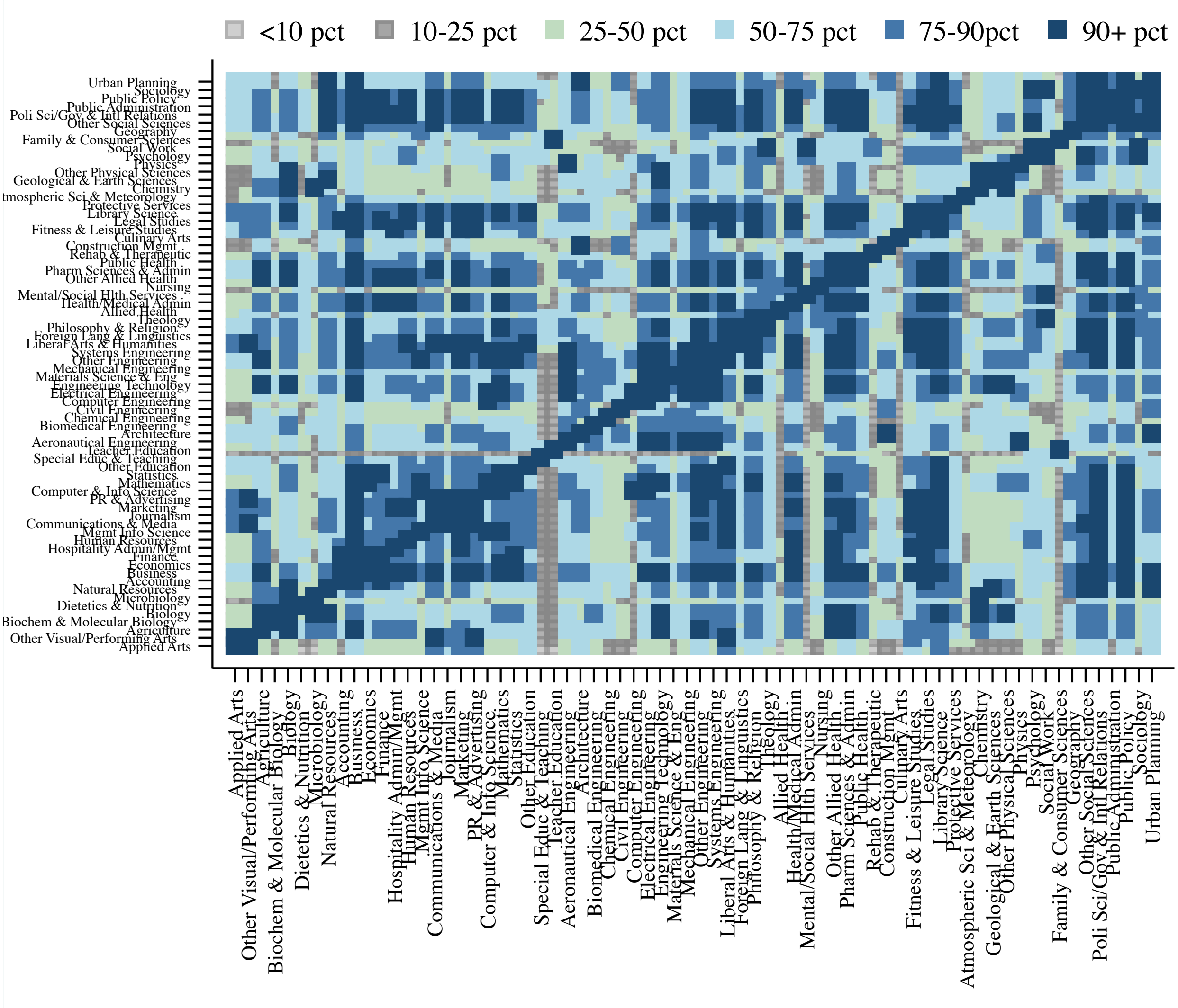


(b) Weighted by Number of Job Ads



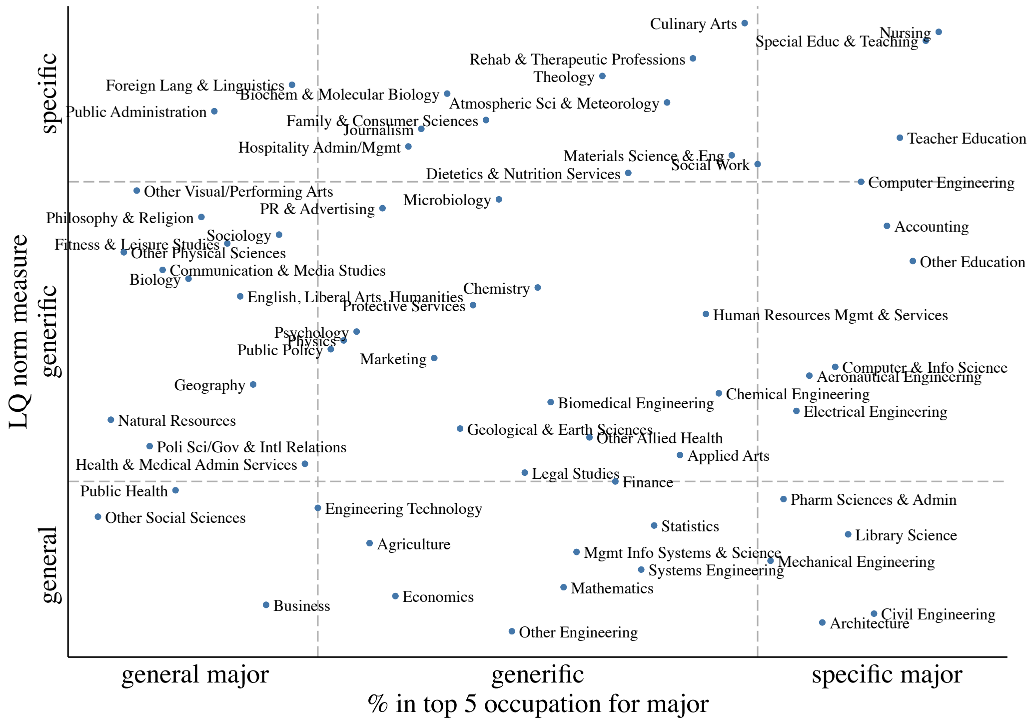
Note: Figure plots the rank of 70 majors using two different measures of skill similarity as measured in the Burning Glass data. The y-axis plots the rank of majors from general (rank=1) to specific (rank=70). Majors are ranked according to the LQ-norm measure which is the sum of the absolute deviations of the major's 11 LQs, from 1: $\sum_{s=1}^{11}[abs(LQ_{s,m}-1)]$. The X-axis plots the rank of each major using the cosine similarity measure constructed using the 9000 most frequent skills (see text for details). In Panel A, majors are unweighted; in Panel B, the circle size represents the number of job postings for the major.

Figure 2.8: Skill Similarity between Each Pair of Majors



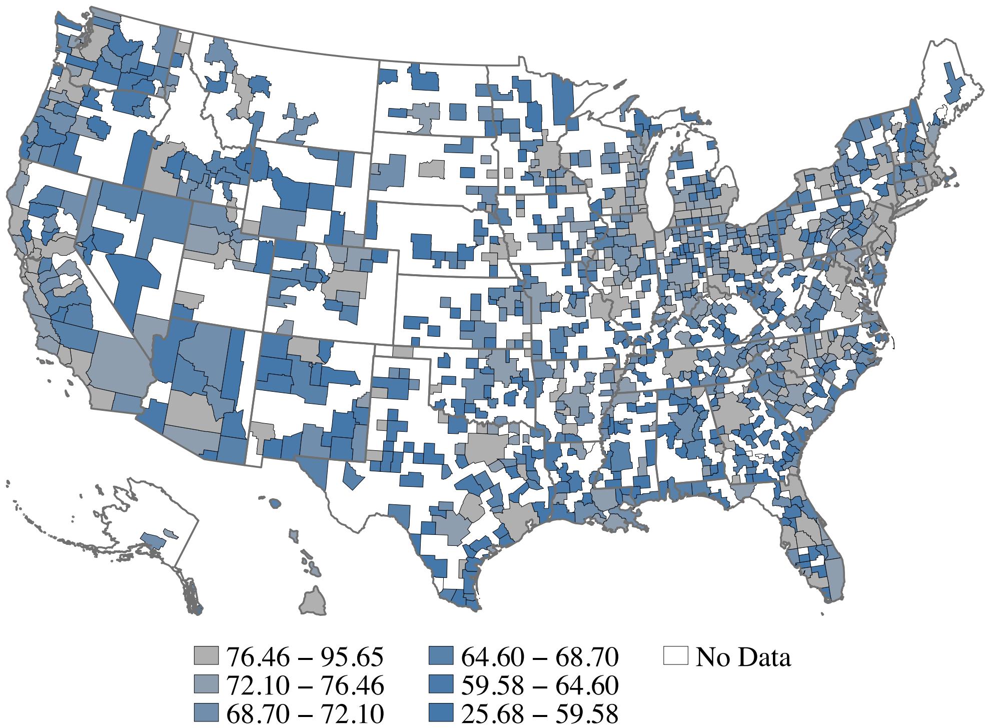
Note: Figures plots a measure of similarity between all possible pairs of majors. Similarity between majors is calculated using the cosine similarity measure and each major's vector of the 9000 most frequent skills (see text for more detail). Cells are colored according to the unweighted percentiles of the distribution of the similarity measures across all majors. Darker cells represent majors that are more similar in terms of skill demand. Similarity measures at different percentiles of the distribution are: 0-10th percentile (similarity = 0-0.21). 10th-25th percentile (0.21-0.40), 25th-50th percentile (0.40-0.51), 50th-75th percentile (0.51-0.63), 75th-90th percentile (0.63-0.72) and above the 90th percentile (0.72-1.00).

Figure 2.9: Skill Similarity Index vs. Occupational Measure of Concentration



Note: Figure plots the rank of 70 majors using two different measures of skill similarity and skill demand as measured in the Burning Glass data. The y-axis plots the rank of majors from general (rank=1) to specific (rank=70). Majors are ranked according to LQ-norm measure which is the sum of the absolute deviations of the major's 11 LQs from 1: $\sum_{s=1}^{11}[abs(LQ_{s,m}-1)]$. The X-axis plots the rank of each major using the percentage of recent college graduates found in the five most frequent occupations for the major as measured in the American Community Survey (ACS). Majors with a lower percentage of recent graduates in the top five occupations are considered more general. Correlation = 0.469 (weighted by number of job postings) and 0.004 (unweighted).

Figure 2.10: Variation in Cognitive Skill Demand Across MSAs, Business Majors



Note: Figure plots the percentage of a metro or micro statistical area's Business major job postings that require cognitive skills as measured in the Burning Glass data.

Table 2.1: Occupational Distribution by Sample

	Samples				
Restriction	(1) All	(2)	(3)	(4)	(5) Analysis
At least 1 Skill		X	X	X	X
Educ = 16			X	X	X
Has a Major				X	X
In Metro CBSAs					X
Count of unique ads	153,031,199	148,000,000	35,938,213	19,519,480	18,471,199
Count of unique ad-major				32,847,216	31,153,536
% of original sample remaining	100	96.71	23.48	12.76	12.07
Mean Experience Level			3.391	3.649	3.682
Occupation (Soc2): % of Postings					
Management (11)	11.70	11.92	22.22	21.93	21.84
Business/Financial (13)	6.64	6.80	14.30	14.82	15.02
Computer/Math (15)	11.54	11.85	22.13	25.23	25.83
Architecture/Engineering (17)	3.15	3.22	6.70	9.50	9.26
Life/Physical/Social Science (19)	1.00	1.03	1.69	2.04	1.97
Community/Social Service (21)	1.09	1.09	1.38	1.40	1.28
Legal (23)	0.85	0.87	0.41	0.25	0.26
Education/Training/Library (25)	2.49	2.52	2.48	1.31	1.25
Arts/Design/Entertainment (27)	2.37	2.42	2.53	2.29	2.32
Healthcare Practitioners (29)	12.27	12.24	7.58	8.21	8.01
Healthcare Support (31)	2.03	2.06	0.01	0.01	0.01
Protective Service (33)	1.00	0.99	0.33	0.22	0.21
Food Prep/Serving (35)	3.38	3.24	0.24	0.23	0.23
Building/Maintenance (37)	1.11	1.11	0.06	0.04	0.04
Personal Care (39)	1.75	1.75	0.27	0.21	0.20
Sales (41)	11.76	12.03	8.20	4.37	4.38
Office/Admin Support (43)	9.96	10.17	4.28	3.02	3.02
Farming/Fishing/Forestry (45)	0.06	0.06	0.02	0.02	0.02
Construction/Extraction (47)	0.97	0.98	0.09	0.11	0.11
Installation/Repair (49)	2.94	3.00	0.31	0.27	0.25
Production (51)	2.45	2.45	0.64	0.56	0.52
Transportation/Material Moving (53)	5.81	4.51	0.14	0.09	0.09
Military (55)	0.07	0.07	0.03	0.02	0.02
Missing (0)	3.61	3.61	3.93	3.84	3.85

Note: Each column is a sample of job postings in the Burning Glass data. Column (1) contains all job postings and Column (5) is the analysis sample. Entries are the percent of job postings in the sample that list each occupation. Occupations are two-digit Standard Occupation Classification (SOC) codes.

Table 2.2: Skill Composite Definition and Examples

Skill	Definition	# skills in top 1000	Top 3 skills	Keywords like Deming and Kahn (2018)
Social	Communicating, persuading, or negotiating with others, which involves adept presentation or exchange of information and perspectives as well as the capacity to accurately infer the motivations of others.	56	Communication Skills, Teamwork /Collaboration, Building Effective Relationships	communication, teamwork, collaboration, negotiation, presentation
People Management	Supervising, motivating, or directing people internal to the business toward defined goals.	43	Staff Management, Leadership, Mentoring	supervisory, leadership, management, mentoring, staff
Cognitive	Applying analytic, logical, quantitative or qualitative reasoning, evaluation, or critical thinking to understand patterns and solve problems.	168	Problem Solving, Research, Creativity	solving, research, analy-, thinking, math, statistics, decision
Writing	Composing, drafting, and editing of books, papers, reports, releases, scripts and other text-based documents; excludes underwriting (which is cognitive).	20	Writing, Written Communication, Editing	writing
Customer Service/Client management	Attracting, soliciting, maintaining, and retaining clients and customers; most forms of sales fall here if there is a personal contact (sales engineering or analysis is cognitive).	110	Customer Service, Sales, Customer Contact	customer, sales, client, patient
Organization	Organizing, planning, managing, and expediting meetings, conferences, events, and other time-sensitive activities; but not logistics or supply chains (which are project management); ability to balance and prioritize among competing demands, apportion work, and meet deadlines.	37	Planning, Organizational Skills, Detail-Oriented	organized, detail-oriented, multitasking, time management, meeting deadlines, energetic

Note: Table lists the author-created 11 mutually exclusive skill composite categories based on the 1,000 most frequent skills in the Burning Glass data. Columns list the definitions of the skill composites and the three most frequently listed skills in each category. Final columns lists the phrases and words used to define these categories in Deming and Kahn (2018). See text for details.

Continued: Skill Composite Definition and Examples

Skill	Definition	# skills in top 1000	Top 3 skills	Keywords like Deming and Kahn (2018)
Computer	General computer tasks and knowledge, including MS Office and related frontline computer support; excludes computer engineering, hardware, design, and other specialized tasks.	22	Microsoft Excel, Microsoft Office, Computer Literacy	computer, spreadsheets, microsoft excel, powerpoint, microsoft office, microsoft word
Software	Use or design of any specialized software, as well as any computer hardware design and engineering, and computer security or network management.	233	SQL, Software Development, Oracle	Skill is categorized as software by BGT
Financial	Preparing or auditing payroll, budgets, accounting or tax documents, and financial reports and statements; excludes financial trading (social), financial engineering, or quantitative financial analysis (both cognitive) – the distinction is that the financial composite captures highly prescribed and rules-based activities that are often ancillary to main activities (unless the main activity is auditing/accounting).	84	Budgeting, Accounting, Procurement	budgeting, accounting, finance, cost
Project Management	Orchestrating, overseeing, or directing programs, projects, processes, and operations – the distinction with people and client management is that the emphasis here is not on people, but rather on the substance of the plans and activities executed by people.	111	Project Management, Quality Assurance and Control, Business Process	project management
Other	Highly discipline-specific skills (often in health) or physical skills that do not readily generalize to other tasks	116	Physical Abilities, Retail Industry Knowledge, Repair	

Note: Table lists the author-created 11 mutually exclusive skill composite categories based on the 1,000 most frequent skills in the Burning Glass data. Columns list the definitions of the skill composites and the three most frequently listed skills in each category. Final columns lists the phrases and words used to define these categories in Deming and Kahn (2018). See text for details.

Table 2.3: Share of Ads for Select Majors Indicating Demand for Each Skill Composite

Major	Cognitive	Social	Project Mgmt	Organizational	Software	Customer Service	Computer Service	Financial	Writing	People Mgmt	Communication	Other Skills (> top 1000)	Other Skills (< top 1000)
All postings	80	68	65	58	50	46	42	43	35	33	46	38	78
Journalism	76	90	44	74	34	40	47	21	100	26	51	35	85
Computer & Info Science	82	65	70	50	94	39	27	19	36	29	47	25	84
Teacher Education	60	99	24	57	4	61	22	17	24	34	28	40	51
Mechanical Engineering	94	58	72	51	48	31	38	37	30	25	43	56	84
Foreign Lang & Linguistics	61	90	30	39	23	16	27	15	44	17	28	30	84
English, Liberal Arts	73	84	40	60	26	36	44	26	60	25	44	32	75
Biology	91	61	54	51	24	29	35	26	36	27	41	69	93
Public Administration	75	69	79	70	23	38	43	67	49	55	36	100	76
Economics	100	75	68	64	45	44	60	61	39	30	52	30	79
Sociology	96	76	42	58	14	65	38	26	37	48	34	58	74
Public Health	77	74	98	58	22	48	44	39	44	43	46	53	84
Nursing	47	60	31	49	4	82	23	16	14	36	30	70	62
Accounting	73	61	52	62	35	33	62	92	30	28	46	28	68
Business	78	77	77	65	40	56	51	56	36	43	53	35	75
Minimum	31	43	15	38	1	15	19	11	12	16	20	25	40
Maximum	100	99	100	87	100	84	63	92	100	76	63	100	100
Mean	79	70	56	57	33	42	38	34	38	34	42	49	81
Standard Deviation	15	12	19	10	24	17	12	17	14	12	9	18	12

Note: Entries are the percent of job postings for the major that list each skill as measured in the Burning Glass data. Mean and standard deviation are calculated equally weighting 70 majors; minimum and maximum are across all 70 majors. Communication skills are also included in Social skills.

Table 2.4: Fraction of Variation in Skill Content Explained by Major and Place

	Variation in skill-share explained by...			
	Major	CBSA	Major & CBSA	Unexplained
Cognitive	0.69	0.07	0.74	0.26
Computer	0.58	0.07	0.64	0.36
Customer service	0.75	0.04	0.78	0.22
Financial	0.84	0.03	0.86	0.14
Organizational	0.42	0.07	0.48	0.53
People management	0.64	0.05	0.68	0.32
Project management	0.71	0.05	0.75	0.25
Social	0.64	0.07	0.71	0.29
Communication skills	0.41	0.11	0.52	0.48
Software	0.87	0.07	0.90	0.10
Writing	0.69	0.06	0.73	0.27
Other (top 1000)	0.69	0.06	0.74	0.26
Unclassified (outside top 1000)	0.61	0.07	0.66	0.34

Note: Observation is a major-MSA cell measuring containing the share of ads seeking each skill as measured in the Burning Glass data. We regress skill demand on major fixed effects only, on MSA fixed effects only and on major and MSA fixed effects. Observations are weighted using major-MSA person weight (perwt) from the American Community Survey (ACS). The amount of variation unexplained is equal to 1 minus the r-squared. Communication skills are also included in Social skills.

Table 2.5: Relationship between Skills and MSA-Major Average Earnings

	Panel A. Base Model			Panel B. Robustness			
	log(raw hourly income)			Adjusted income		Unweighted	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cognitive	0.399***	0.223*	-0.00001	0.259**	-0.00793	0.271***	0.0002
	(0.142)	(0.117)	(0.026)	(0.117)	(0.029)	(0.078)	(0.013)
Computer	-0.253**	-0.0658	-0.0540***	-0.0202	-0.0687***	-0.0408	-0.0143
	(0.106)	(0.070)	(0.016)	(0.060)	(0.017)	(0.046)	(0.015)
Customer	0.0809	0.0432	0.0291	0.125	0.0257	-0.03	0.0152
	(0.110)	(0.089)	(0.023)	(0.078)	(0.022)	(0.066)	(0.013)
Financial	0.303***	0.235***	-0.00855	0.158**	-0.0102	0.0506	-0.010
	(0.079)	(0.069)	(0.024)	(0.066)	(0.023)	(0.062)	(0.016)
Organizational	-0.187	-0.269**	-0.00845	-0.258***	-0.0139	-0.176***	-0.0115
	(0.113)	(0.108)	(0.016)	(0.094)	(0.016)	(0.038)	(0.013)
People mgmt	-0.609***	-0.489***	-0.0184	-0.345***	-0.0147	-0.178***	0.00603
	(0.146)	(0.130)	(0.032)	(0.095)	(0.033)	(0.055)	(0.015)
Project mgmt	0.401***	0.375***	0.0206	0.207**	0.00502	0.280***	0.0187
	(0.112)	(0.093)	(0.024)	(0.080)	(0.025)	(0.073)	(0.016)
Social	-0.317**	-0.477***	0.00794	-0.365***	0.0156	-0.193***	0.00396
	(0.146)	(0.119)	(0.019)	(0.104)	(0.019)	(0.051)	(0.016)
Software	0.02	-0.0372	0.018	-0.0955	0.0245	0.0405	0.00346
	(0.115)	(0.101)	(0.023)	(0.085)	(0.024)	(0.060)	(0.018)
Writing	0.000129	-0.0546	-0.00841	-0.0417	0.000973	-0.114***	0.0119
	(0.112)	(0.102)	(0.022)	(0.088)	(0.021)	(0.037)	(0.015)
Other (top 1000)	-0.102	-0.0478	-0.0486*	0.0114	-0.0503*	-0.0333	-0.0312**
	(0.115)	(0.100)	(0.025)	(0.099)	(0.030)	(0.056)	(0.015)
Constant	3.648***	3.908***	3.665***	3.789***	3.668***	3.458***	3.474***
	(0.169)	(0.146)	(0.040)	(0.150)	(0.047)	(0.088)	(0.018)
Observations	22,151	22,151	22,151	22,151	22,151	22,151	22,151
R^2	0.342	0.621	0.870	0.588	0.830	0.228	0.466
Weights	YES	YES	YES	YES	YES	NO	NO
Major FE	NO	NO	YES	NO	YES	NO	YES
MSA FE	NO	YES	YES	YES	YES	YES	YES
F-test (all 11)	17.94***	13.24***	2.863***	8.894***	2.583***	15.829***	2.41**

Note: Variables are the share of job postings in the major-MSA cell that list each skill as measured in the Burning Glass data. Outcome is the log of mean hourly earnings (2019 dollars) among college graduates in each major-MSA cell as measured in the 2009-2018 American Community Survey (ACS). Adjusted income is regression-adjusted for compositional differences across majors. Earnings sample is restricted to full-time, year-round workers who are not enrolled in education at the time of the survey. Observations include all workers aged 25-54 except in columns 8 and 9 which is restricted to ages 23-34. The F-test is a test of joint significance for all skill variables. Standard errors are two-way clustered by MSA and major. Weights are major-MSA person weight (perwt) from the ACS. *** p<0.01, ** p<0.05, * p<0.1.

Continued: Relationship between Skills and MSA-Major Average Earnings

	Panel B. Robustness Continued			
	Wages age<35		Ads experience 0-2	
	(8)	(9)	(10)	(11)
Cognitive	0.0554	-0.007	0.224**	0.0176
	(0.105)	(0.025)	(0.098)	(0.022)
Computer	-0.130**	-0.0693***	-0.00843	-0.0384**
	(0.060)	(0.019)	(0.063)	(0.015)
Customer	0.144*	0.0201	0.0275	0.0195
	(0.081)	(0.024)	(0.078)	(0.020)
Financial	0.188***	0.00877	0.212***	-0.0125
	(0.067)	(0.022)	(0.063)	(0.016)
Organizational	-0.282***	-0.00354	-0.243***	-0.0058
	(0.106)	(0.022)	(0.087)	(0.012)
People management	-0.278***	0.00614	-0.437***	-0.0401
	(0.093)	(0.025)	(0.115)	(0.026)
Project management	0.324***	0.00384	0.312***	0.00827
	(0.091)	(0.024)	(0.085)	(0.018)
Social	-0.442***	-0.00115	-0.431***	-0.00665
	(0.113)	(0.019)	(0.098)	(0.014)
Software	0.115	-0.0054	-0.0211	0.0311
	(0.096)	(0.022)	(0.095)	(0.021)
Writing	-0.102	-0.0249*	-0.0813	-0.000185
	(0.095)	(0.015)	(0.083)	(0.013)
Other skills (top 1000)	-0.0556	-0.0482*	-0.0573	-0.0342*
	(0.088)	(0.029)	(0.082)	(0.020)
Constant	3.632***	3.377***	3.878***	3.660***
	(0.142)	(0.041)	(0.121)	(0.028)
Observations	19,480	19,480	21,614	21,614
R^2	0.587	0.806	0.616	0.871
Weights	YES	YES	YES	YES
Major FE	NO	YES	NO	YES
MSA FE	YES	YES	YES	YES
F-test (all 11 skills)	15.266***	2.389**	12.532***	2.91***

Note: Variables are the share of job postings in the major-MSA cell that list each skill as measured in the Burning Glass data. Outcome is the log of mean hourly earnings (2019 dollars) among college graduates in each major-MSA cell as measured in the 2009-2018 American Community Survey (ACS). Adjusted income is regression-adjusted for compositional differences across majors. Earnings sample is restricted to full-time, year-round workers who are not enrolled in education at the time of the survey. Observations include all workers aged 25-54 except in columns 8 and 9 which is restricted to ages 23-34. The F-test is a test of joint significance for all skill variables. Standard errors are two-way clustered by MSA and major. Weights are major-MSA person weight (perwt) from the ACS. *** p<0.01, ** p<0.05, * p<0.1.

CHAPTER III

Decomposing Earnings Differences between College Majors

3.1 Introduction

Graduates with different college majors enter the labor market with very different types of human capital (Arcidiacono, 2004; Zafar, 2013; Hastings et al., 2013; Kirkeboen et al., 2016). As employers perceive different majors as having different skills (Hemelt et al., 2021a) and jobs can vary substantially in the skills they require (Acemoglu and Autor, 2011; Deming and Kahn, 2018), skill differences between majors could lead to dramatically different career paths post graduation. To the extent that different types of work are associated with different pay – for example, pay differences across occupations (Goldin, 2014) and firms (Abowd et al., 1999; Card et al., 2013; Barth et al., 2016; Song et al., 2019) – differences between majors in typical job attributes could account for a significant portion of the differences in mean earnings between majors (Altonji et al., 2012, 2016a).

This paper asks to what extent are differences in mean earnings between majors accounted for by mean differences in typical job attributes (i.e. sorting). To answer this, I use data from the National Survey of College Graduates (NSCG) to measure differences in mean earnings between majors in two specifications. The first specification is a baseline model with basic demographic controls and the second also includes extensive labor market controls that measure a worker's job attributes. The baseline model measures earnings differences that occur *between* and *within* different types of work, whereas the latter isolates earnings differences to those that occur *within* different types of work. In each model, between-major differences in average earnings are measured using the standard deviation of the major fixed effects, and each major is weighted by survey weighted-employment.

I find that over half of the baseline differences in mean earnings between majors is accounted for (statistically) by mean differences between majors in job attributes. Job attributes include occupation, primary and secondary work tasks, supervising others, graduate education, region of employment and employer-ownership structure, -size and -age. At baseline the standard deviation of the major fixed effects is .168 but decreases to .082 once job attributes are controlled for. The remaining earnings differences between majors reflects earnings variation that occurs among graduates performing similar types of work.[1]

I next use decomposition methods to partition the total decrease in between-major differences in average earnings into parts due to mean differences in each type of job attribute. I use the Gelbach (2016) decomposition as the results are invariant to the order in which explanatory controls are added to the full model, and it allows me to assess the empirical relevance of all job attributes simultaneously. An individual covariate X will account for differences in mean earnings between majors if there are mean differences in the variable across college majors (i.e. sorting) and the variable is conditionally correlated with earnings.

I find that, perhaps unsurprisingly, occupation accounts for the largest share of the explained between-majors earnings inequality (37%). However, differences between majors in the work activities performed on the job, including supervising others, explain an additional 18% of the differences in mean earnings between majors. Most surprising is the substantial role employer characteristics play. College majors are not equally distributed across small or large firms, nor are they equally likely to work at for-profit businesses or the federal government. Accounting for these differences explains an additional 40% of the earnings differences.

This paper contributes to a large literature that seeks to understand the sources of earnings differences between college majors (Altonji et al., 2012; Hershbein and Kearney, 2014; Hastings et al., 2013; Kirkeboen et al., 2016).[2] Past work shows that college majors are concentrated in particular occupations (Ransom and Phipps, 2017; Altonji et al., 2012) and that there are earnings premiums associated with working in an occupation that is related to a worker's primary field of study (Robst, 2007; Nordin et al., 2010; Kinsler and Pavan, 2015). I too find that occupation of employment is a key determinate of earnings differences between majors.

In contrast to past studies which primarily focus on differences between majors in the occupation of employment, I also investigate if the typical job attributes of college majors differ along other dimensions, and whether this matters for earnings variation. One notable

[1]Note that this is not exactly equivalent to isolating earnings variation into a particular job or firm as in done in other papers that look at earnings differences due to sorting across establishments. For example, see Barth et al. (2018).

[2]This analysis also relates to a literature that uses decomposition methods to understand the determinants of race- and gender-wage gaps (see Neal and Johnson (1996), Lang and Manove (2011) and Goldin (2014) among many others).

exception is Bleemer and Mehta (2022), who use a regression discontinuity design and find that half of the wage return for economics majors is explained by working in higher-paying industries. I (descriptively) build upon their work by leveraging rich data in the NSCG to examine a wide variety of job attributes – including job tasks, whether a worker supervises others and employer characteristics – and earnings inequality for the entire spectrum of college majors. I find that there is extensive variation in the tasks workers perform on the job (e.g. research or management) even within occupation (Spitz-Oener, 2006; Autor and Handel, 2013; Deming and Kahn, 2018). This variation is empirically important, as job-level variables explain, conditional on occupation, an additional 20% of the between-major differences in average earnings.

I also document that college majors tend to work for different types of employers. Similar to prior research, these results suggest that education plays a salient role in allocating workers across heterogenous firms (Engbom and Moser, 2017; Cardoso et al., 2018; Ost et al., 2019; Huneeus et al., 2021b).[3] Moreover, I find that the sorting of college majors across employer attributes is as important as occupation in accounting for between-majors earnings inequality. This result relates to numerous recent studies that stress the relationship between earnings inequality and between-firm pay differences (Abowd et al., 1999; Card et al., 2013; Barth et al., 2016; Song et al., 2019).

The rest of this paper proceeds as follows. Section 3.2 describes the data, Section 3.3 measures earnings inequality between majors. Section 3.4 discusses the decomposition and which covariates matter for explaining between-majors earning inequality. Section 3.5 concludes.

3.2 Data

The main data source is the 2003 and 2010-2019 waves of the National Survey of College Graduates (NSCG). The sampling frame for the 2003 wave was the 2010 Decennial Census, and from 2010-2019 was respondents to the 2009-2017 American Community Survey (ACS) who had a bachelor's degree at the time of survey. Starting in 2010, the NSCG follow a rotating panel design and individuals can be surveyed up to four times, but I use the NSCG as repeated cross-sections. While the NSCG surveys college graduates in all academic disciplines,

[3]These studies use firm fixed effects and employer-employee linked data. Cardoso et al. (2018) and Engbom and Moser (2017) find that up to one quarter of the pay premiums for higher degrees can be explained by between-firm pay differences and the sorting of more educated individuals to higher paying firms by including job and establishment fixed effects. Similarly, Barth et al. (2018) find that the observed and unobserved measures of employers mediate the effects of years of schooling on earnings.

there is a particular focus on graduates in the science and engineering workforce.[4] As a result the unweighted distribution of graduates in the NSCG may be somewhat skewed, but Appendix Table C.1 shows that applying survey weights makes the distribution of college majors and demographic characteristics in the analysis sample similar to those in a nationally representative sample.[5]

In the main analysis I restrict attention to observations for individuals aged 23 to 65 at the time of survey, who graduated with their first bachelor's degree in year 1965 or later, in the United States, and between the ages of 20-26. I also restrict attention to workers who are employed in the United States, who work full-time at their primary job (at least 35 hours a week) and full-year (at least 41 weeks per year).[6] Finally, I exclude observations with a missing undergraduate major, year of bachelor's degree graduation or occupation codes. The final analysis sample consists of 250,265 observations.

The main dependent variable is log annual earnings measured by the annualized salary at the primary employer in a given reference week. All earnings data are converted to 2012 values using the Personal Consumption Expenditures (PCE) inflator. Following Altonji et al. (2014) and Altonji and Zhong (2021), I exclude annual earnings that are less to $5,000 and top code earnings at 400,000.[7]

College major is categorized using the individual's primary field of study for their first bachelor's degree. College majors are classified using 61 detailed college major categories based on the American Community Survey (ACS) college major codes and created by Hemelt et al. (2021a). I also aggregate a few of their codes to adjust for the NSCG college major codes (see Martin (2022) for more details).

I next describe the set of independent variables that will be used throughout the analysis. In all analyses the variables are divided into two sets of controls corresponding to two regression models: the *baseline* and *full* model. The baseline model covariates are capture pre-college attributes and time-specific variables. The full model covariates collectively describe a worker's employment attributes (e.g. occupation or employer type), and are meant to control for the differential sorting of college majors across types of work.

[4] Science and Engineering fields of study include: Computer and Math sciences; Biological, agricultural, and environmental life sciences; Physical sciences (Physics, chemistry, geosciences); Social Sciences (Psychology, economics, political science); and Engineering.

[5] I use sample weights that are adjusted to accommodate the different sizes of the survey waves. See Appendix Section C.2 for more details.

[6] My choice of 41 weeks per year follows Altonji and Zhong (2021) who impose this restriction to accommodate the employment arrangements of many teachers.

[7] These restriction affect roughly 1% and less than .5% of observations, respectively.

3.2.1 Baseline Model Variables

Demographic variables include indicators for female, black and hispanic. One key limitation of the NSCG is that it lacks any measure of pre-college aptitude (e.g. standardized test score like the SAT). I instead measure parents' education, a dummy equal to one if at least one of the respondent's parents has a bachelor's degree, which serves as a crude proxy for pre-college or unobservable ability and to address selection into major based on pre-college test scores (Arcidiacono, 2004).For some analyses, I measure a worker's type of undergraduate institution using the 1994 Carnegie Classification and following Hersch (2013).[8] The groups include (1) private Research I and private Research II universities, (2) private Liberal Arts I colleges, (3) public Research I universities, and (4) colleges in the remaining Carnegie classifications, excluding specialized institutions.[9] Finally, I include survey year fixed effects, a quartic in years since first bachelor's degree graduation (as a measure of potential experience), ten-year graduation cohort fixed effects, and fixed effects for region of high school attendance.

3.2.2 Full Model Variables

The full model include dummies for the geographic division of a respondent's employer to capture differential access to labor markets that is a function of college major. I include controls for graduate degrees with separate indicators for master's degree (MS MA MBA), professional degree (JD LLB MD DDS DVM etc) and doctoral degree (PhD DSc EdD).

Employer Attributes The NSCG has several variables that measure employer characteristics including -size, -age and -ownership structure. I measure employer size using a series of 8 indicators based on the NSCG categorical variable: < 10 employees, 11-24, 25-99, 100-499, 50-999, 1000-4999, 5000-24999 and 25000+ employees. Based on the specific wording of the question, the NSCG variable more closely resembles the concept of firm employment rather than establishment employment. Employer age is measured by an indicator for whether the employer was established as a new business within the past 5 years. Employer ownership (which is similar to class of worker in the ACS) is measured using a series of 8 indicators based on the NSCG categorical variable: pre-college institutions (i.e. K-12 schools) and two-year colleges, four-year colleges/university research institute/medical school, for-profit

[8]Hersch (2013) compares the overlap of schools in each category of the Barron's Profile of American Colleges (a commonly used measure of institutional selectivity) and the 1994 Carnegie Classification. Institutions are then divided so that the share of schools rated as most or highly competitive is significantly different between groups. See Table 1 of Hersch (2013) for details.

[9]I also create a fifth group which includes graduates from specialized institutions, non-US institutions, and those that are missing information on Carnegie classification.

business or industry; non-profit business or industry, self-employed (both incorporated and not incorporated), state or local government and federal government (including U.S. military).

Occupation & Job Tasks I classify the occupation of each worker's primary job into 83 categories. See Appendix Section C.2 for a list of the occupation codes.

The NSCG also includes questions about the work activities individuals perform at their primary job. These questions enable an analysis of job task variation at the individual level and within occupation, which past evidence suggests is substantial (Spitz-Oener, 2006; Autor and Handel, 2013; Deming and Kahn, 2018). This variation could explain differences in mean earnings between majors to the extent that two workers with different college majors are assigned different job tasks within occupation (due to skill differences). Respondents indicate all of the work tasks they spend at least 10% of their week performing at their primary job.[10] Respondents are also asked to indicate which work activities they spend the most (and second most) hours on during a typical week which are called their primary and secondary work activities, respectively. Possible work activities include: accounting, basic research, applied research, development, design, computer applications, employee relations, managing, production and operations, professional services, sales, quality or productivity management and teaching. See Appendix Section C.2 for definitions of these work tasks.

Two exercises suggest that there is meaningful variation in job tasks not only across occupations, but also within occupations. To illustrate this, Appendix Table C.2 presents the occupations with the highest occurrence of each primary work task, and Appendix Table C.3 presents the most commonly reported primary work activities by occupation. The most common occupations performing project & people management (teaching) are Mangers (Teachers), and similarly, 52.2% (91.7%) of Managers (Primary and Secondary Teachers) report that their primary work activity is project & people management (teaching). Second, a variance decomposition, displayed in Table 3.1, show that a non-trivial part of variation in the work activities occurs within occupations rather than between occupations.[11] While for some work activities like Teaching, between-occupation variation comprises around 80% of the total variation in the task, for most other tasks the between-occupation differences

[10] On average individuals select roughly four different work activities that they typically perform on their principal job. Only 1% of respondents in the analysis sample indicate 0 tasks and an additional 1% indicate more than 10 tasks (i.e. tasks that total more than 100% of their work week). There is minimal variation across majors in the number of tasks reported.

[11] To do this I employ an identity taken from Barth et al. (2016): $V\big(task_{ijk}\big) = V_w + V_b = V\big(task_{ijk} - E[task_{ijk}]\big) + V\big(E[task_{ijk}]\big)$, where $task_{ijk}$ is the indicator for whether worker i performs task k in occupation j, $E[task_{ijt}]$ is the mean of the task j indicator for workers in occupation j, V_w is the within component of the variance, and V_b is the between component. For each work activity k, I use the individual level data and regress the indicator $task_{ijk}$ on j occupation dummies. I calculate the variance of the residuals to approximate $V\big(task_{ijk} - E[task_{ijk}]\big)$ and the variance of the occupation fixed effects to approximate $V\big(E[task_{ijl}]\big)$.

account for less than 50% of the total variation. This is true whether or not the task is one a worker spends at least 10% of their week performing or if it is the task the worker spends the most hours per week performing (i.e their primary task).

Supervisory Work The NSCG asks whether individuals supervise others in their primary job, which is defined as recommending or initiating personnel actions such as hiring, firing, evaluating or promoting others.[12] Appendix Figure C.1 shows that the percent of individuals supervising others has a quadratic relationship with years since graduation and this pattern is present both between and within occupation. This suggests that increases in supervisory work are capturing changes in job levels, both within and across occupations, that occur as workers accumulate experience (Gibbons and Waldman, 1999, 2006).

3.3 Differences in Mean Earnings between Majors

3.3.1 Empirical Approach

I first measure the earnings differences between majors that reflect the combined impact of sorting across various job attributes (described in the previous section) and earnings differences within work types. I then estimate earnings differences that don't reflect sorting by adding in the full model covariates which controls for mean earnings differences across work types. In both cases, between-major differences in average earnings are measured using the standard deviation of the model's estimated major fixed effects, where each major is weighted by survey weighted-employment.

To calculate the baseline earnings differences between majors, I regress annual log earnings on fixed effects for each of the 61 majors and a limited set of control variables:

$$log(earn_{imt}) = {\beta_0}^{base} + \sum_m {\beta_m}^{base}(D_m) + {\beta_x}^{base} X_{imt} + \epsilon \tag{3.3.1}$$

where $log(earn_{imt})$ is the log of annual earnings for individual i with major m measured in year t, the variable D_m is a dummy for each of the majors m and X_{imt} is the vector of baseline covariates. The estimated coefficient $\hat{\beta}_m^{base}$ on each major fixed effect measures the earnings advantage of major m over the omitted major of Other Education residualized on the baseline variables X_{imt}.

[12] Respondents are also explicitly told teachers should not count students in their response to this question.

To account for sorting across job attributes, I add in the full set of labor market controls:

$$log(earn_{imt}) = \beta_0{}^{full} + \sum_m \beta_m{}^{full}(D_m) + \beta_x{}^{full} X_{imt} + \beta_z{}^{full}(Z_{imt}) + \epsilon \tag{3.3.2}$$

where $log(earn_{imt})$, D_m and X_{imt} are as before, and Z_{imt} is the set of labor market controls (see Section 3.2 for a description of these variables). The coefficient $\hat{\beta}_m^{full}$ now measures the earnings advantage of major m over the omitted major of Other Education residualized on both the baseline variables (X_{imt}) and the labor market variables (Z_{imt}).

For a single major, a reduction in the magnitude of the major fixed effect between Equation (3.3.1) and (3.3.2) (i.e. $|\hat{\beta}_m^{base}| > |\hat{\beta}_m^{full}| > 0$) indicates that accounting for mean differences in Z decreases the major's earnings premium. That is, a portion of the earnings advantage of major m over Other Education majors is accounted for by differences in the types of work major m performs, and the the associated earnings premiums. A reduction in the between-major differences in average earnings occurs if the standard deviation of the major fixed effects decreases from the baseline to the full model. In particular, if $\sigma(\hat{\beta}_m^{base}) > \sigma(\hat{\beta}_m^{full})$, then accounting for differences in the types of work performed between majors can *explain* (in an accounting, not causal, sense) earnings inequality between majors.[13]

3.3.2 Results

For the vast majority of majors, the major's earnings premium over Other Education majors *decreases* when accounting for differences in the labor market attributes Z. Figure 3.1a and 3.1b plots the distribution of the baseline major fixed effects ($\hat{\beta}_m^{base}$) from Equation (3.3.1) and the full model major fixed effects ($\hat{\beta}_m^{full}$) from Equation (3.3.2).[14] As Other Education is one of the lowest earnings majors, most majors have a positive earnings premium, averaging around 29% and ranging from -4% among Philosophy, Religion & Theology majors to 69% for Management Information Systems & Sciences majors. The distribution of major fixed effects in the full model is visibly compressed relative to the baseline model and the average major now earns only 9.5% more than Other Education. Figure 3.1c more clearly illustrates the changes in the earnings premiums by plotting the base model fixed effects against the full model fixed effects for each major. The point for all majors falls below the 45 degree line – indicating a decrease in the major's earnings premium – and for the median major, the earnings premium in the full model is 36% of the baseline premium.

[13]Note that while the magnitude of the major fixed effect $\hat{\beta_m}$ for major m is sensitive to the choice of omitted major the variance of the major fixed effects, and thus the estimated extent of earnings inequality between majors, is not.

[14]See Appendix Table C.4 for the major fixed effects and values of $\hat{b}_{m,std}$ for all majors and both models.

For most majors accounting for differences in work types doesn't substantially change the major's relative position in the overall earnings distribution. To illustrate this, major fixed effects from each specification are standardized by subtracting the mean and variance and majors are weighted by survey-weighted employment: for the base model = $\left[\hat{\beta}_m^{base} - \overline{\hat{\beta}_m^{base}}\right] / \left[\sigma(\hat{\beta}_m^{base})\right]$ and for the full model=$\left[\hat{\beta}_m^{full} - \overline{\hat{\beta}_m^{full}}\right]\left[\sigma(\hat{\beta}_m^{full})\right]$. In Figure 3.1d, the points for most majors are scattered around the 45 degree line, indicating little change in the standardized values across specifications. In addition, when majors are ranked using their fixed effects from the base model and separately for the full model, the correlation between the ranks is .94 ($R^2 = .88$) and the median major moves only 2 places in the distribution of 61 majors.[15]

Finally, changes in between-major differences in average earnings from the baseline to the full model can be summarized by the standard deviation of the major fixed effects. At baseline it is equal to 16.9% but in the full model it decreases to 8.2%. Thus, controlling for earnings differences that are due to differences between college majors in mean job attributes, reduces the baseline differences in mean earnings between majors by 52%. The remaining 48% is unexplained by the differential sorting of college majors and mean earnings differences across occupation, employer attributes, and job tasks. It is important to note that the inclusion of other variables measuring job attributes, including industry, or other types of job tasks not capturing in the NSCG, may further explain earnings differences between majors.

3.4 Decomposing Earnings Differences

3.4.1 Empirical Approach

Accounting for mean differences between majors in job attributes halves between-majors earnings inequality. This is due to the addition of numerous variables including occupation, primary and secondary work tasks, supervisory work, graduate education and employer region, sector, size and age. Which of these covariates are most empirically salient to the overall decrease? In this section, I employ decomposition methods commonly used in the gender- and race-wage gap literature (Neal and Johnson, 1996; Lang and Manove, 2011; Goldin, 2014) to answer this question in three steps. I first construct a single variable that indexes college major and measures baseline earnings inequality between majors. I then measure the change in the coefficient on this variable between the baseline and full model. Finally, I use the Gelbach (2016) decomposition to partition the total change in the coefficient between models

[15] For 75% of the majors, the change in rank is only plus or minus 2 positions. There are a few majors with large changes including Library Science which improved from rank 52 to 34, Legal Studies moved down from 30^{th} to 47^{th} and Nutritional Sciences moved down from 42^{th} to 58^{th}.

into parts due to each variable that is in the full but not the baseline model.

3.4.1.1 Standardized Major Fixed Effect

Decompositions of earnings gaps typically involve analyzing which variables contribute most to changes in the coefficient on a dummy for group membership (e.g. female indicator). While this approach is technically feasible for considering earnings differences between majors, it is undesirable for a few reasons. First, there are 61 college majors and thus 60 dummy variables to be decomposed. Second, the coefficient on each dummy variable measures the earnings premium of major m relative to some arbitrarily chosen omitted major, which is often not the salient earnings gap. Finally, this approach doesn't measure changes in the distribution of earnings across *all* majors.

Given these complications, I instead construct a single variable whose value varies by college major and apply the decomposition to the coefficient on that variable. Following Altonji et al. (2014) and Altonji et al. (2016b), I construct a standardized major fixed effect which I refer to as $\hat{b}_{m,std}$. Specifically, I standardize the major fixed effects that I estimated using the baseline model (Equation (3.3.1)) to be mean zero and standard deviation one:

$$\hat{b}_{m,std} = \frac{\hat{\beta}_m - \overline{\bar{\hat{\beta}}_m}}{\sigma(\hat{\beta}_m)}, \qquad \bar{\hat{\beta}}_m = \frac{\sum_m \hat{\beta}_m}{m} \qquad \sigma(\hat{\beta}_m) = \sqrt{\frac{\sum_m (\hat{\beta}_m - \bar{\hat{\beta}}_m)^2}{m-1}} \tag{3.4.1}$$

where I drop the superscript *base* for ease of notation. The mean and standard deviation of the 61 major fixed effects are given by $\bar{\hat{\beta}}_m$ and $\sigma(\hat{\beta}_m)$ and are calculated using each major's survey-weighted employment. The standardization collapses the 61 estimated baseline major fixed effects into a single standardized variable, which measures the earnings premium of that major, in standard deviation units, relative the average major.

I then replace the 61 major dummies in Equation (3.3.1) with the single standardized major fixed effect:

$$log(earn_{imt}) = {\alpha_0}^{base} + {\alpha_1}^{base}\hat{b}_{m,std} + {\alpha_x}^{base} X_{imt} + \epsilon \tag{3.4.2}$$

The baseline model now has a single variable that measures college major, and more importantly, the estimated coefficient on that variable ($\hat{\alpha}_1^{base}$) has an economically meaningful interpretation. In particular, the coefficient $\hat{\alpha}_1^{base}$ on $\hat{b}_{m,std}$ estimated in Equation (3.4.2) is mechanically equivalent to the standard deviation of the baseline major fixed effects estimated in Equation (3.3.1): $\hat{\alpha_1}^{base} = \sigma(\hat{\beta}_m^{base})$. See Appendix C.3 for details.[16] Thus, the coefficient

[16] As both Equation (3.3.1) and (3.4.2) measure the same covariance between the baseline independent variables X_{imt} and the dependent variable $log(earn_{imt})$, the estimated coefficients on X_{imt} in both models

on $\hat{b}_{m,std}$ corresponds to the baseline level of between-major differences in average earnings

I then estimate the full model, in which I replace the 61 major dummies in Equation (3.3.2) with the variable $\hat{b}_{m,std}$ that was constructed using the base model fixed effects:

$$log(earn_{imt}) = {\alpha_0}^{full} + {\alpha_1}^{full}(\hat{b}_{m,std}) + {\alpha_x}^{full} X_{imt} + {\alpha_z}^{full}(Z_{imt}) + \epsilon \tag{3.4.3}$$

I measure the total change in the coefficient on $\hat{b}_{m,std}$ between models: $\hat{\delta} = \hat{\alpha}_1^{base} - \hat{\alpha}_1^{full}$. If the change is positive, then adding in the controls Z to account for the sorting of majors across work types reduces the baseline earnings inequality between majors.[17] Given the economic interpretation of $\hat{\alpha}_1$, a decomposition of a decrease in the coefficient $\hat{\alpha}_1$ is equivalent to a decomposition of a decrease in the between-major differences in average earnings between models.

Measurement Error of $\hat{\beta}_m$ Each major fixed effect $\hat{\beta}_m$ estimates the earnings premium of major m with error due to sampling variability. This will lead to an overstatement of between-majors earnings inequality, and will introduce an errors-in-variables problem created by using $\hat{b}_{m,std} = \frac{\hat{\beta}_m - \overline{\hat{\beta}_m}}{\sigma(\hat{\beta}_m)}$ as an independent variable.

Specifically, suppose that the true major fixed effect for major m is β_m and the observed major fixed effect is $\hat{\beta}_m = \beta_m + e$ where e is the estimation error. The variance of the estimated fixed effects is $\text{Var}(\hat{\beta}_m) = \text{Var}(\beta_m) + \text{Var}(e)$, which is larger than the true variance.[18] In turn, the variable $\hat{b}_{m,std}$ will be understated by a factor of $\sqrt{\text{Var}(\beta_m)}/\sqrt{\text{Var}(\hat{\beta}_m)}$, and the values of $\hat{b}_{m,std}$ should be multiplied by a factor of $\sqrt{\text{Var}(\hat{\beta}_m)}/\sqrt{\text{Var}(\beta_m)}$. As before, $\sqrt{\text{Var}(\hat{\beta}_m)}$ is measured by $\sigma(\hat{\beta}_m)$. An estimate of the true variance $\text{Var}(\beta_m)$ can be obtained by subtracting from $\text{Var}(\hat{\beta}_m)$ an estimate of the mean error variance. Following Jacob and Lefgren (2008), I estimate the mean error variance using the squared standard errors on the estimated major fixed effects: $\overline{[\text{se}_m(\hat{\beta}_m)]^2}$, which I find is equal to .0002185. As this is a relatively minor correction compared to the estimated variance of the major fixed effects (.1687), I proceed with using the uncorrected version of $\hat{b}_{m,std}$.[19]

will be equivalent: $\hat{\beta}_x^{base} = \hat{\alpha}_x^{base}$. This will only be true if exactly the same set of covariates used to estimate $\hat{\beta}_m^{base}$ in Equation (3.3.1) are also included in regression Equation (3.4.2). See Appendix C.3 for details

[17]Note that while mechanically it is true that ${\hat{\alpha}_1}^{base} = \sigma(\hat{\beta}_m^{base})$, the same statement is not true for the full model coefficient and full model fixed effects ${\hat{\alpha}_1}^{full} \neq \sigma(\hat{\beta}_m^{full})$. If a model with only X_1 is used to estimate $\hat{\beta}_m$ and create $\hat{b}_{m,std}$, but the coefficient $\hat{\alpha}_1$ on the same $\hat{b}_{m,std}$ is estimated in a model with variables X_1 and X_2, then the coefficient on $\hat{\alpha}_1$ doesn't exactly equal $\sigma(\hat{\beta}_m$ but is very similar.

[18]This exposition closely follows that as presented in Jacob and Lefgren (2005) and Jacob and Lefgren (2008).

[19]In future work I will estimate the sensitivity of the results to this correction.

3.4.1.2 Gelbach Decomposition

I use the Gelbach (2016) decomposition to partition the total change in the coefficient on $\hat{b}_{m,std}$, $\hat{\delta} = \hat{\alpha}_1^{base} - \hat{\alpha}_1^{full}$, into parts due to each job attribute variable added to the full model. The Gelbach (2016) decomposition is a desirable method as it can be applied to a coefficient on a non-binary variable (e.g. unlike the Oaxaca-Blinder decomposition) and the results are invariant to the order in which explanatory controls Z_k are added to the model. Gelbach (2016) shows that the change in $\hat{\alpha}_m^{base}$ due to the addition of a variable Z_k depends on the sequence in which covariates are added to the model because both, one, the correlation between Z_k and the other covariates Z_j $(k \neq j)$ and, two, the conditional correlation between Z_k and earnings, depend on which other Z_j have already been added to the model.Given there is usually not a natural sequence in which to add in covariates, Gelbach (2016) proposes a decomposition based in the omitted variable bias formula which simultaneously accounts for the role of all variables.

Briefly, the intuition is that if the baseline model with labor market controls is the "naive" model and the full model with labor market controls is the "true" model, then the difference between the coefficients in the naive (base) and true (full) model can be expressed as an omitted variable bias. The population omitted variable bias from excluding Z when estimating α is plim $\hat{\alpha} - \alpha = (X'X)^{-1}X'Z\beta_Z$. In this application the equivalent expression is $\hat{\alpha}_m^{base} - \hat{\alpha}_m^{full} = (\hat{b}'_{m,std}\hat{b}_{m,std})^{-1}\hat{b}'_{m,std}Z\hat{\beta}_Z$ in which $\hat{\alpha}_m^{full}$ is substituted for α, $\hat{\alpha}_m^{base}$ for $\hat{\alpha}$ and $\hat{b}_{m,std}$ for X.

The linearity of the omitted variable bias formula implies that if Z consists of k different covariates then the total change in $\hat{\alpha}_1$ can be subdivided into parts due to each covariate Z_k:

$$\hat{\delta} = \hat{\alpha}_m^{base} - \hat{\alpha}_m^{full} = \sum_k \hat{\delta}_k = \sum_k (\hat{b}'_{m,std}\hat{b}_{m,std})^{-1}\hat{b}'_{m,std}Z_k\hat{\beta}_k \tag{3.4.4}$$

Equation (3.4.4) enables the calculation of $\hat{\delta}_k/\hat{\delta}$ which is the share of the total change in the coefficient $\hat{\alpha}_1$ that is due to the addition of Z_k to the regression model. In practice, I group the individual covariates into groups g (e.g. occupations) and measure the extent to which different covariate groups change the estimated coefficient: $\hat{\delta}_g/\hat{\delta} = \sum_{k\in g}\hat{\delta}_k/\hat{\delta}$ for all $k \in g$. For each covariate group g, the decomposition statistically quantifies the degree to which between-major differences in average earnings would change if the distribution of college majors across the covariates in g would be equivalent to the distribution among the entire sample of all college graduates (Cardoso et al., 2018).[20]

[20] In contrast to the standard Oaxaca-Blinder decomposition (Blinder, 1973; Oaxaca, 1973) there is only one full model and not a model separately estimated for each group. Thus all individuals, regardless of group membership, face the same coefficient β_k for covariate X_k. In the standard Oaxaca-Blinder decomposition

In what case does an individual covariate or a group of covariates contribute to the change in the coefficient? To see this, examine the two terms in Equation (3.4.4). The first term, $(\hat{b}'_{m,std}\hat{b}_{m,std})^{-1}\hat{b}'_{m,std}Z_k$, is the coefficient on $\hat{b}_{m,std}$ from an auxiliary linear model projecting Z_k on $\hat{b}_{m,std}$ and the baseline controls. If the coefficient on $\hat{b}_{m,std}$ is non-zero, this indicates that college majors sort across the particular work attribute Z_k. The second term $\hat{\beta}_k$, is the coefficient on Z_k from the full earnings model and is equal to the correlation between Z_k and earnings conditional all other variables in the full model. Thus, for an individual regressor Z_k to account for a non-zero portion of the change in $\hat{\alpha}_1$ between the base and the full model, there must be (1) sorting across the variable by college major (e.g. a correlation between Z_k and $\hat{b}_{m,std}$) and (2) the variable must be correlated with earnings conditional on all other covariates (e.g. $\hat{\beta}_k \neq 0$). In the case of a groups of covariates, the linearity of Equation (3.4.4) implies that the impact of the individual covariates in the group must not cancel each other out. Moreover, in the case of covariate groups, the above two conditions are necessary but not sufficient for the coefficient on $\hat{b}_{m,std}$ to be impacted by controlling for the variables $k \in g$ (Gelbach, 2016).

3.4.2 Results

I first estimate the major fixed effects with Equation (3.3.1) and then standardize them using the (survey-weighted employment) mean and standard deviation. The full distribution of the major fixed effects was discussed previously and illustrated in Figure 3.1a, but Figure 3.2 shows the distribution of detailed majors with high- and low-values of $\hat{b}_{m,std}$ across broad curriculum-based categories of majors. Both Computer Science & Engineering majors and Business & Economics majors have many detailed majors with high standardized fixed effects ($\hat{b}_{m,std} > 1$). All Education majors have low values ($\hat{b}_{m,std} < -1$). Communications & Marketing majors and Health majors tend to fall in the middle of the distribution ($-1 < \hat{b}_{m,std} < 0$ and $0 < \hat{b}_{m,std} < 1$). All Humanities majors and Other majors have values below the average major ($\hat{b}_{m,std} < 0$).

Panel A of Table 3.2 displays the coefficients $\hat{\alpha}_1$ on $\hat{b}_{m,std}$ in the baseline and full model estimated using Equation (3.4.2) and (3.4.3). At baseline, the coefficient is .1687 (p<.01) and adding in all the covariates decreases the coefficient to .0769, which is a decrease of 0.0918 (p< .01) log points and represents a 54% decrease. Thus, controlling for differential sorting of majors with low- and high- values of the standardized major fixed effect across different types of work more than halves the measure of between-major differences in average earnings

earnings gaps reflect both differences in the mean of Z_k across groups *and* differences in the return to Z_k across groups (i.e β_k is group-specific). The Gelbach decomposition is nested in the Oaxaca-Blinder. For details see Gelbach (2016).

to 7.6%.[21]

Panel B partitions the total change in the coefficient on $\hat{b}_{m,std}$ into parts due to each of the covariate groups. The first column reports $\hat{\delta}_g$, the log points accounted for by the covariate group g conditional on all of the other covariates simultaneously. The third column presents $100 \times \hat{\delta}_g/\hat{\delta}$, which is the percent of the total change accounted for by the covariate group g. Figure 3.3 also plots $100 \times \hat{\delta}_g/\hat{\delta}$ to summarize the relative importance of each covariate group.

The most important variables in accounting for earnings inequality between majors include occupation, employer characteristics (size and ownership) and job-level variables (work tasks and job level). Together these variables account for over 90% of the decrease in the coefficient on $\hat{b}_{m,std}$. Specifically, occupation accounts for roughly 37.5% of the total decrease, primary and secondary work activities explain an additional 13%, supervising others explains an additional 5%, employer ownership structure explains 26%, and employer size explains an additional 13%.[22] The remaining 6% is explained by differences in graduate education, employer size, and employer region, but for all of these covariate the figures are not statistically different from zero.

Robustness to Alternative Baseline Models Results in Column (2) of Table 3.4 show that the pattern of results are qualitatively similar across perturbations of which variables are included in the baseline model. In general, the baseline level of between-major differences in average earnings changes, but the relative importance of job attributes are stable. First, if the baseline model used to create $\hat{b}_{m,std}$ includes no control variables (i.e. based on raw differences in mean earnings between majors), then the baseline differences in mean earnings between majors increase (.1963 compared to .1687). The total decrease in the coefficient from the baseline to the full model is also larger (61%) because demographic characteristics positively account for between-major differences in average earnings, but are treated as a factor used to account for earnings differences (and account for 16% of the total decrease). Second, results in Column (4) reveal little differences between majors when survey weights aren't used. Third, baseline between-major differences in average earnings decrease when I include type of undergraduate institution, in the baseline model. Appendix Figure C.2 plots, for four groups of the variable $\hat{b}_{m,std}$, the distribution of individuals across each institutional

[21]Unsurprisingly, Table 3.2 shows that the coefficient on $\hat{b}_{m,std}$ is exactly equal to .1687, the standard deviation of the major fixed effects.

[22]Note that the importance of occupation is not sensitive to whether I used detailed or broad occupation codes. As the NSCG survey has a focus on STEM workers, the occupation codes vary in the level of detail across broad groups of occupations. For example, there are more occupation codes for Computer Occupations than for Business and Financial Occupations, an occupation that is less associated with science, technology, engineering, and mathematics (STEM) workers. Column (3) of Table 3.4 shows that results using an aggregation of the detailed codes into roughly 20 broad occupation codes following Altonji and Zhong (2021) are very similar.

category. The figure illustrates that the distribution of institution type varies across majors. For instance, majors with the highest values of baseline earnings ($\hat{b}_{m,std} > 1$) are more likely to have attended private Research I-II universities but less likely to have attended Liberal Arts I colleges. Results in Column (3) show that the baseline between-major differences in average earnings are lower when institutional category is included in the baseline model (.156 compared to .1687 without).

3.4.3 Detailed Decomposition

I analyze further why particular work attributes empirically matter in the decomposition. Recall that Section 3.4.1.2 shows that for an individual regressor Z_k to account for a non-zero portion of the change in the coefficient between the base and the full model, two things must be true. First, there must be sorting across the variable Z_k by college major conditional on the *baseline* controls, and, second, the variable must be correlated with earnings conditional on the *full* model controls. I investigate, one, whether the coefficient on $\hat{b}_{m,std}$ is non-zero in a regression of Z_k on the baseline controls and, two, whether the coefficient $\hat{\beta}_k$ on Z_k is non-zero in a regression of log earnings on the full model controls.

Occupation & Job Tasks Variation occupation accounts for .0345 log points (p<.01) or roughly 37.5% of the total differences in mean earnings between majors that is explained statistically by the full model covariates. Three pieces of analysis illustrate why this is. First, Figure 3.4a illustrates that college majors are not equally distributed across occupations. For each occupation, the figure plots the coefficient on $\hat{b}_{m,std}$ from a regression of the occupation dummy on $\hat{b}_{m,std}$ and the baseline covariates. For many occupations, the coefficient on $\hat{b}_{m,std}$ is significant, indicating that there is a statistically significant difference in the probability of working in the occupation between a major with $\hat{b}_{m,std} = 0$ and a major with $\hat{b}_{m,std} = 1$. Second, occupation is correlated with earnings conditional on all other covariates in the model. This is illustrated by Table 3.3 which performs a F-test of joint significance on coefficients for the occupation dummies in the full model (F=9797, p<.01). Finally, majors with high values of $\hat{b}_{m,std}$ tend to be overrepresented in occupations that are positively correlated with earnings. Figure 3.4b illustrates this by plotting the coefficients on each occupation dummy from the full model against the coefficients on $\hat{b}_{m,std}$ from Figure 3.4a.

The primary and secondary work activities individuals perform on the job explain an additional 13% of the between-majors earnings variation. Recall that the between and within occupation variance analysis in Table 3.1 revealed that a non-trivial part of variation in the job tasks occurs within occupations. Appendix Table C.5 shows that for the vast majority of work activities, the amount of variation that is not captured by occupation and major is substantial – including 48% for professional services, 66% for accounting, 98% for quality management.[23] In addition, primary and secondary work activities are correlated with earnings conditional on all other covariates in the model: the F-test of joint significance on coefficients in each group results in an F-statistics of 36.59 (p<.01) and 23.79 (p<.01), respectively. Finally, around a fourth of the total job-level impact results from variation between majors in the propensity to supervise others on the job.

Employer Characteristics Employer characteristics account for roughly 39.5% of the explained differences in average earnings between majors. Of this, around one third is due to employer size and the remaining two thirds is due to employer ownership structure. The impact of controlling for employer age (a dummy for whether or not the employer has existed for at least 5 years) plays a negligible and statistically insignificant role.

Figure 3.5 illustrates that college majors are not equally distributed across employer characteristics. For each employer size and ownership category, the figure plots the coefficient on $\hat{b}_{m,std}$ from a separate regression of a dummy for the category on $\hat{b}_{m,std}$ and the baseline covariates. Majors with a value of $\hat{b}_{m,std}$ ($\hat{b}_{m,std}$=1) are 11 percentage points much likely than the average major ($\hat{b}_{m,std}$=0) to work in for-profit business or industries (p<.01) and slightly more likely to work for the federal government (.8 percentage points, p<.05). These majors, however, are 8.5 percentage points (p<.05) less likely to work for 2-year colleges or pre-college institutions. Majors with higher values of $\hat{b}_{m,std}$ are also more likely to work for larger employers. This sorting are similar to previous studies which have found that workers with different education levels (Engbom and Moser, 2017; Cardoso et al., 2018) and college majors (Ost et al., 2019; Huneeus et al., 2021b) are not evening distributed across heterogenous firms. Both employer ownership and employer size are correlated with earnings conditional on all other covariates in the model: separate tests of joint significance on the coefficients for employer size and ownership structure categories results in an F-statistics of 277.4 (p<.01) and 66.43 (p<.01), respectively.

Graduate Education Variation between majors in graduation attainment plays little role in explaining the between-major differences in mean earnings (4.5%, p=0.369). This is not because graduate attainment is uncorrelated with earnings in the analysis sample; the test of joint significance on the graduate school coefficients in the full model is (F=112, p<.01). However, the coefficients on $\hat{b}_{m,std}$ from separate regressions of dummies for a Master's, Professional or Doctorate degree on $\hat{b}_{m,std}$ and the baseline covariates are not statistically

[23] One unsurprising exception is teaching, for which major and occupation together explain 77% of the variation.

different from zero. Thus, graduate school attainment is associated with a conditional earnings premium, but there are not large differences in attainment between majors with different values of $\hat{b}_{m,std}$. Given large variation in graduate attainment across undergraduate college major, and differences across graduate degrees and fields in earnings effects (Altonji and Zhong, 2021), graduate education could play a larger role in explaining earnings differences between majors using an alternative categorization of majors.

Working Full-Time or Full-Year The main analysis sample only includes individuals that work at least 35 hours per week (full-time) and 41 weeks per year (full-year). As a result, the main analysis excludes variation in hours and weeks worked as an explanation for earnings differences across majors. Roughly 12% of the (weighted) sample works less than 35 hours per week and 9% works less than 41 weeks per year. As majors may differ in the propensity to work full-time or full-year, I run the main analysis including all employed individuals irrespective of work intensity.

Results in Column (5) of Table 3.4 illustrate that the baseline earnings inequality between majors is more extensive among all employed workers than among only full-time full-year workers (.184 compared to .168). In addition, the decrease in the coefficient on $\hat{b}_{m,std}$ is larger. In particular, when all labor market covariates in the full model as well as indicators indicators for full-time (hours>35/week) and full-year (weeks>41/year) are added to the model, the coefficient on $\hat{b}_{m,std}$ decreases to .068, which is a decrease of .1153 or 63%. As with full-time full-year workers, the main drivers of the reduction in between-majors earnings inequality include occupation, employer sector, employer size and job tasks.

The overall larger reduction in between-majors earnings differences, relative to the main specification for full-time full-year workers, is due to the importance of labor supply in accounting for earnings differences. Differences between majors in the propensity to work full-time or full-year account for roughly 15% of the explained earnings differences.[24] Majors with high values of $\hat{b}_{m,std}$ are more likely to work full-time and full-year, both of which are associated with positive conditional earnings premiums.

3.5 Conclusion

With large increases in college enrollment over the last 40 years, researchers are increasingly interested in understanding why labor market outcomes vary substantially with a worker's type of college education. One crucial way in which graduates differ is college major, which is

[24] Of these, about two-thirds is due to the full-time indicator and the remaining third is due to the full-year indicator.

related to the pre-labor market skills of graduates (Arcidiacono, 2004; Zafar, 2013; Hastings et al., 2013; Kirkeboen et al., 2016). Skill differences between majors are likely one reason that college majors are not equally dispersed across occupations (Ransom and Phipps, 2017; Altonji et al., 2012). However, as there is extensive variation in the tasks workers perform on the job (Spitz-Oener, 2006; Autor and Handel, 2013; Deming and Kahn, 2018), and there are productivity differences across firms (Syverson, 2011), the typical job performed by a college major likely also differs along additional dimensions. To the extent that different types of work are associated with different pay – for example, pay differences across occupations (Goldin, 2014) and firms (Abowd et al., 1999; Card et al., 2013; Barth et al., 2016; Song et al., 2019) – differences between majors in typical job attributes could account for a significant portion of the differences in mean earnings between majors (Altonji et al., 2012, 2016a)

Using data from the National Survey of College Graduates (NSCG), I find that over half of the baseline between-major differences in mean earnings can be accounted for by mean differences between majors in typical job attributes. I then use the Gelbach (2016) decomposition to understand which job attributes – including occupation, primary and secondary work tasks, supervising others, employer ownership structure and employer size – are most empirically salient in accounting for earnings differences between majors. A particular job attribute will account for earnings differences between majors if there are mean differences in the variable across college majors (i.e. sorting) and the variable is conditionally correlated with earnings. I find that over 90% of the decrease in between-majors differences in average earnings is due to the differential sorting of college majors across occupation, employer characteristics (size and ownership) and job-level variables (work tasks and job level). Unsurprisingly, occupation accounts for the largest share of the total explained between-majors earnings differences (37%). However, differences between majors in the work activities performed on the job, including supervising others, explain an *additional* 18% of the between-majors earnings variation.

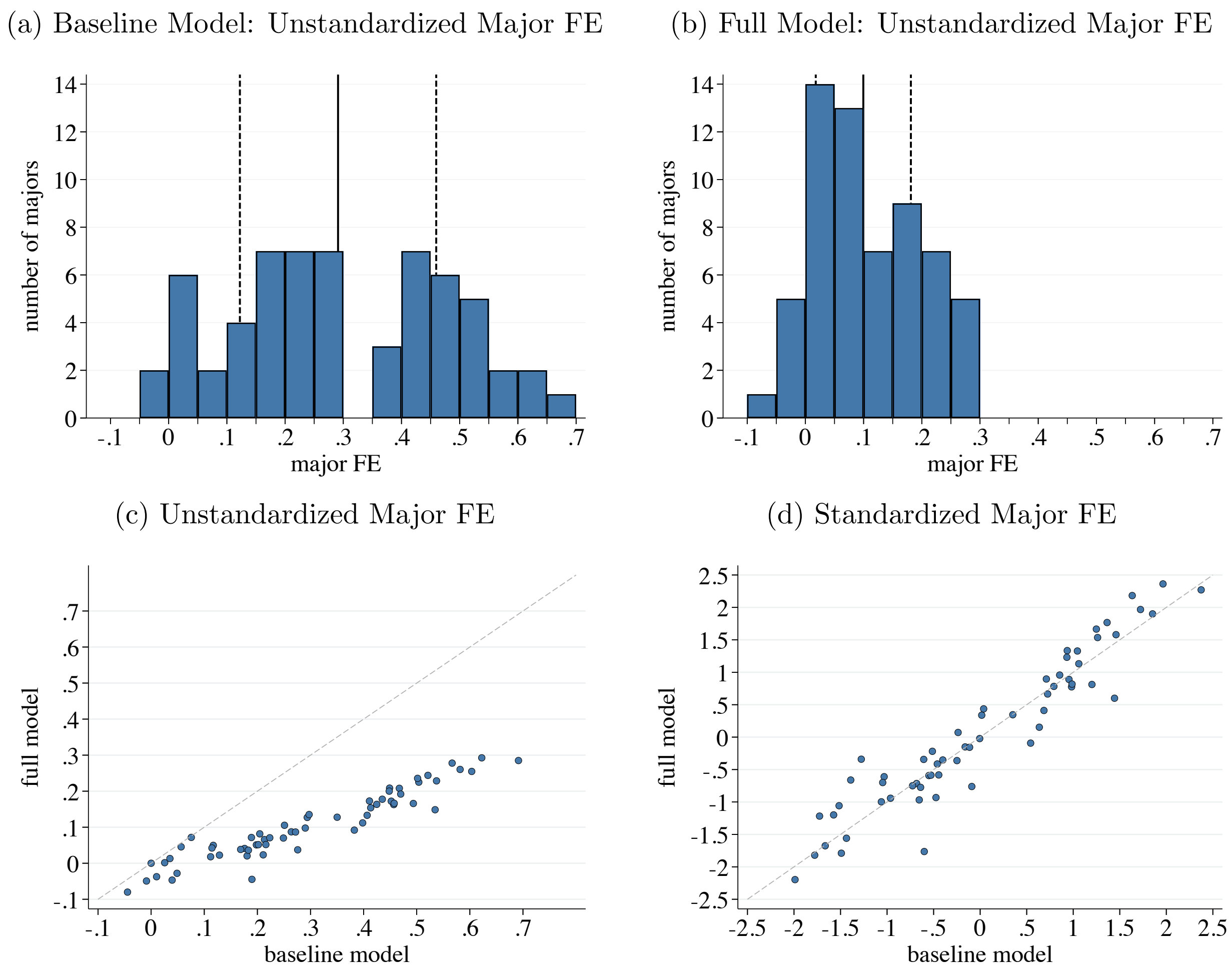
Most surprising is the substantial role employer characteristics, including ownership structure and size, play in explaining between-majors differences in average earnings. College majors are not equally distributed across small or large firms, nor are they equally likely to work at for-profit businesses and the federal government, and controlling for this sorting accounts for 40% of the between-majors earnings inequality. This suggests that an investigation of the relationship between college major and employer attributes (e.g the recruitment strategies of employers) would further improve an understanding of the mean earnings differences between majors.

Table 3.1: Variance Decomposition of Work Activities Between & Within Occupations

	Panel (a): ≥10% of hrs/wk			Panel (b): primary activity		
job task	between	within	total	between	within	total
accounting	28%	72%	0.23	32%	68%	0.07
applied research	8%	92%	0.20	11%	89%	0.03
basic research	5%	95%	0.16	8%	92%	0.01
computer applications	24%	76%	0.17	30%	70%	0.05
development	18%	82%	0.18	12%	88%	0.04
design	8%	92%	0.20	5%	95%	0.03
employee relations	13%	87%	0.22	18%	82%	0.03
project & people mgmt	13%	87%	0.12	14%	86%	0.03
production	36%	64%	0.23	52%	48%	0.16
quality mgmt	12%	88%	0.23	19%	81%	0.16
sales	8%	92%	0.21	2%	98%	0.02
professional services	25%	75%	0.24	34%	66%	0.11
teaching	32%	68%	0.21	77%	23%	0.09

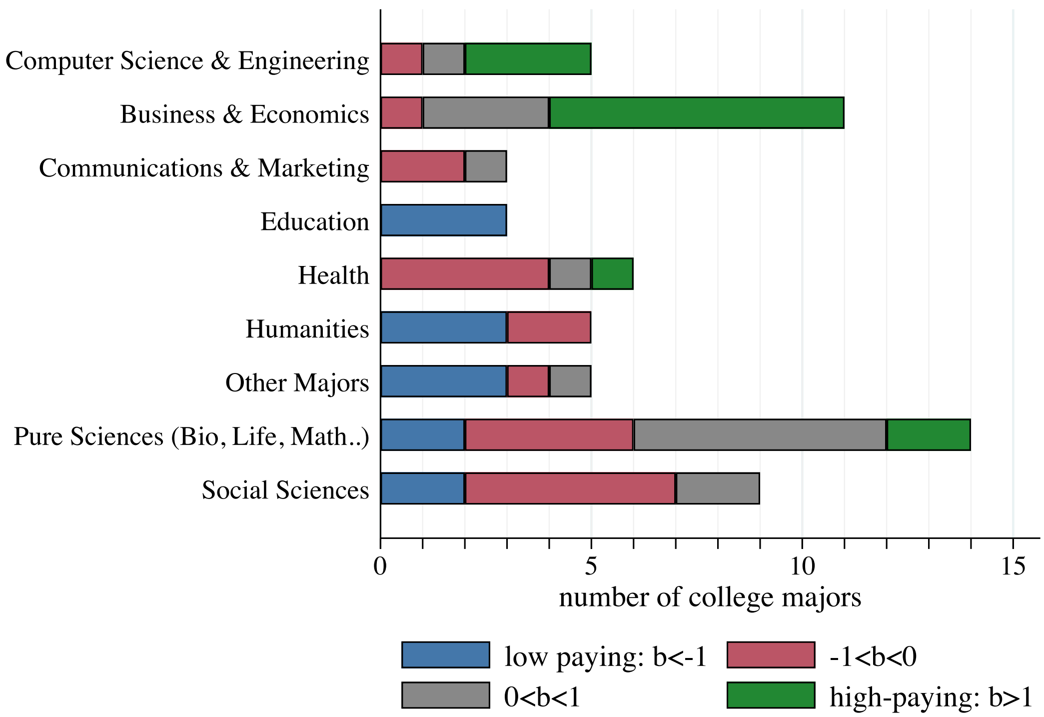
Note: Table presents results of a variance decomposition. Columns labeled *between* (*within*) give the percent of the total variation in the job task (work activity) that occurs between (*within*) occupations. Outcomes in Panel (a) are an indicator for whether the work performs the job tasks at least 10% of hours per week at their primary job. Outcomes in Panel (b) are an indicator for which job tasks the work spends the most hours performing. Results use the variance accounting identity from Barth et al. (2016): $V\big(task_{ijk}\big) = V_w + V_b = V\big(task_{ijk} - E[task_{ijk}]\big) + V\big(E[task_{ijk}]\big)$, where $task_{ijk}$ is the indicator for whether worker i performs task k in occupation j, $E[task_{ijt}]$ is the mean of the task j indicator for workers in occupation j, V_w is the within component of the variance, and V_b is the between component. For each outcome and work activity k, I use the individual level data and regress the indicator $task_{ijk}$ on j occupation dummies. I calculate the variance of the residuals to approximate $V\big(task_{ijk} - E[task_{ijk}]\big)$ and the variance of the occupation fixed effects to approximate $V\big(E[task_{ijl}]\big)$. Observations are weighted using NSCG survey weights. Data source is the 2003-2019 NSCG. N= 250,265.

Figure 3.1: Distribution of Residualized Major Fixed Effects ($\hat{\beta}_m$)



Note: Figure plots the distribution of major fixed effects from two different regressions. Estimates in Panel (a) are from a regression of log annual earnings on 61 major fixed effects and survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, and an indicator for parental college education. Panel (b) are from a regression of log annual earnings on 61 major fixed effects and the controls from Panel (a) and controls for occupation, graduate education, employment region, employer size, employer age, employer ownership structure, thirteen indicators for the type of primary and secondary job tasks (work activity) and an indicator for whether the worker supervises others. In each figure, the solid line is the average of the major fixed effects and the dashed lines represent one standard deviation above or below the mean. The mean and standard deviation are calculated using survey-weighted employment in each major. Panel (c) plots, for each major, the major fixed effect in the baseline and full model. Panel (d) plots, for each major, the *standardized* major fixed effect in the baseline and full model. For each model, standardized fixed effects are calculated using the mean and standard deviation of the 61 major fixed effects (with survey-weighted employment). Data source is the 2003-2019 NSCG. N= 250,265 in each regression.

Figure 3.2: Subject Field Categories: Distribution of Low- and High-Paying Majors



Note: Figure plots the number of college majors in each college major subject field that fall into one of four categories based on the standardized major fixed effect. The standardized major fixed effect is calculated by regressing log annual earnings on 61 college major dummies and survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, an indicator for parental college education, high school region and 10-year graduation cohort fixed effects.. The major fixed effects are then standardized fixed effects are calculated using the mean and standard deviation of the 61 major fixed effects (with survey-weighted employment). The 61 majors are divided into four categories based on the standardized major fixed effect $\hat{b}_{m,std}$ (b in the figure): $\hat{b}_{m,std} < -1$, $-1 < \hat{b}_{m,std} < 0$, $0 < \hat{b}_{m,std} < 1$ and $\hat{b}_{m,std} > 1$. Data source is the 2003-2019 NSCG. N=250,265.

Table 3.2: Gelbach Decomposition of Standardized Major Fixed Effect

Panel A: Change in Primary Coefficient

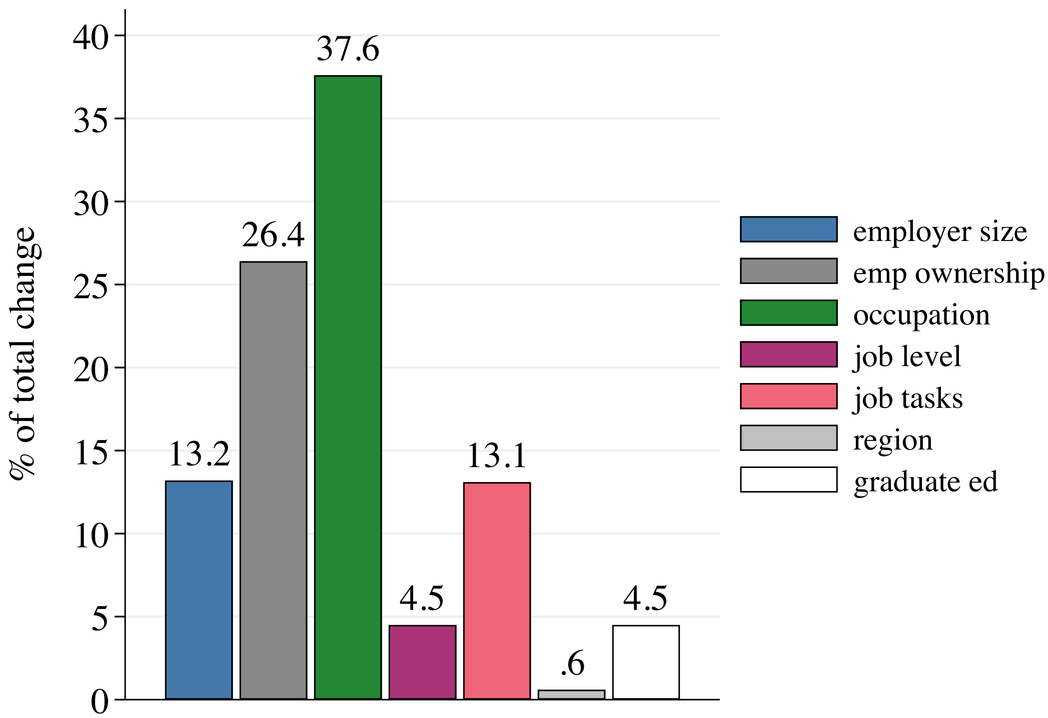
	coefficient	se
α_1^{base}: coeff on $b_{m,std}$ in baseline model	.1687***	(.0011)
α_1^{full}: coeff on $b_{m,std}$ in full model	.0769***	(.0042)
$\hat{\delta} = \alpha_1^{base} - \alpha_1^{full}$ total change	.0918***	(.0041)

Panel B: Total Change Partitioned Across Variable Groups

	Covariate in Model		Results		
	baseline	full	$\hat{\delta}_g$	se	$100\hat{\delta}_g/\hat{\delta}$
baseline covariates					
demographics	Yes	Yes			
survey year	Yes	Yes			
years since graduation	Yes	Yes			
occupation	No	Yes	.0345***	(.0050)	37.5%
graduate education	No	Yes	.0041	(.0049)	4.50%
employer region	No	Yes	.0006	(.0006)	.623%
employer characteristics:					
size	No	Yes	.0122***	(.0033)	13.2%
age	No	Yes	.0000	(.0000)	.004%
ownership	No	Yes	.0242***	(.0043)	26.3%
job level data					
primary work activities	No	Yes	.0098***	(.0028)	10.6%
secondary work activities	No	Yes	.0023***	(.0006)	2.45%
job level (supervisor)	No	Yes	.0042***	(.0015)	4.54%
Observations	250,265	250,265			
Adjusted R-squared	.269	.492			

Note: Table presents results from the Gelbach (2016) decomposition. Entries for $\hat{\delta}_g$ are the log points accounted for by the covariate group g conditional on all of the other covariates simultaneously. Entries in column $100 \times \hat{\delta}_g/\hat{\delta}$ give the percent of the total change accounted for by the covariate group g, where $\sum_g \hat{\delta}_g = \hat{\delta}$. The baseline model is a regression of log annual earnings on the standardized major fixed effect ($b_{m,std}$), survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, and an indicator for parental college education. The full model adds in occupation FE, dummies for master's professional degree and Ph.D degrees, employment region, employer size, an indicator for employer age greater than 5 years, employer sector, thirteen indicators for the type of primary and secondary work activity and an indicator for whether the worker directly supervises individuals in their primary job. Standard errors are clustered by major. Decomposition performed using the stata command *b1x2*. Each regression uses weights equal to survey-weighted employment. Data source is the 2003-2019 NSCG.

Figure 3.3: Percent of Earnings Gap Explained by Covariate Types



Note: Figure plots results from the Gelbach (2016) decomposition and presented in Table 3.2. For each covariate group g, the figure plots $100 \times \hat{\delta}_g/\hat{\delta}$, which the percent of the total change in the coefficient on $b_{m,std}$ accounted for by the covariate group g. Data source is the 2003-2019 NSCG.

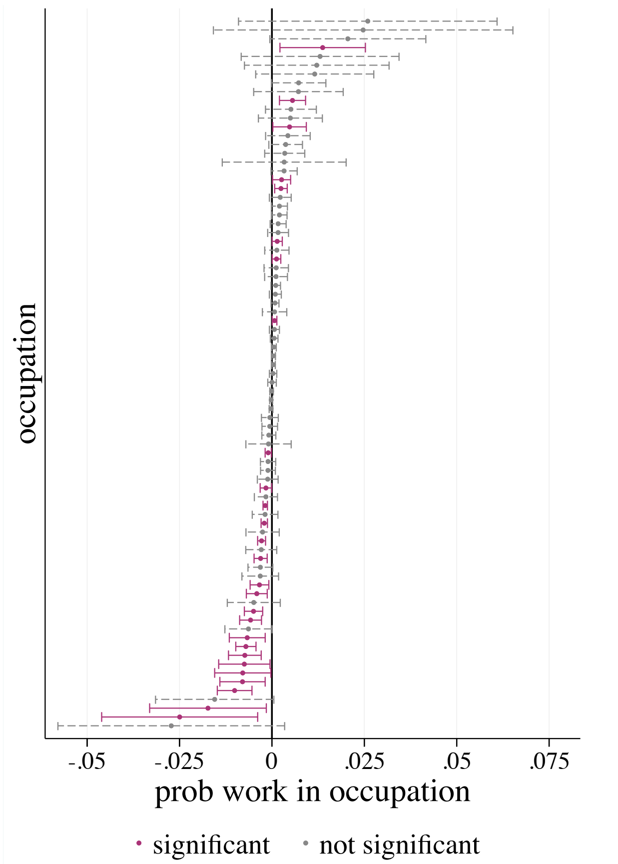
Table 3.3: ANOVA Analysis and F-tests by Covariate Group

	Panel A: ANOVA			**Panel B: F-test**	
	model SS	R^2	%Δ model SS	F-stat	P-value
model without:					
region	46118	0.49	-0.01	93.91	0.00
occupation	46120	0.49	0.00	9796.37	0.00
graduate degree	44824	0.48	-2.81	112.30	0.00
employer characteristics	42622	0.46	-7.59	225.47	0.00
age	46119	0.49	-0.01	0.18	0.67
size	46121	0.49	0.00	277.36	0.00
ownership	45979	0.49	-0.31	66.43	0.00
job level data	44107	0.47	-4.37	109.54	0.00
supervise others	45737	0.49	-0.83	200.62	0.00
work tasks	45501	0.49	-1.34	53.68	0.00
primary tasks	45689	0.49	-0.94	36.59	0.00
secondary tasks	45842	0.49	-0.61	23.79	0.00
job level or employer data	40595	0.43	-11.98		

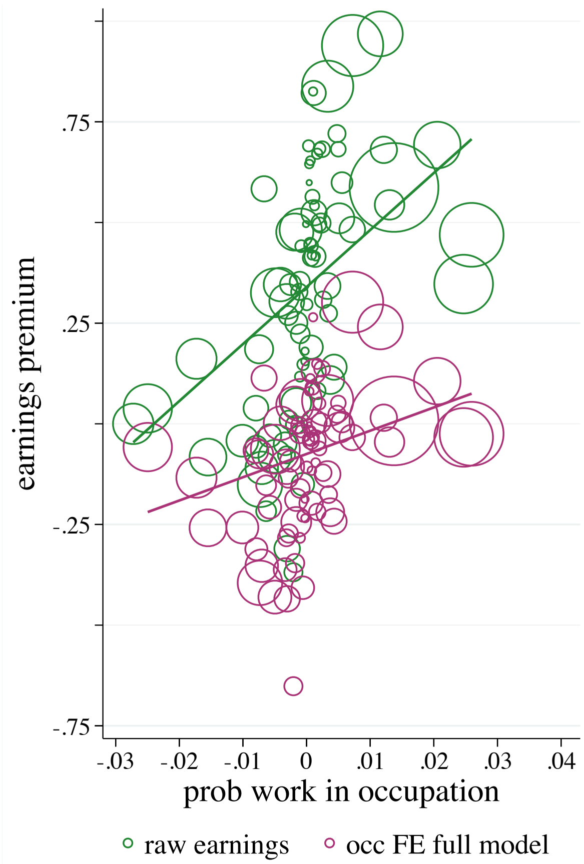
Note: Panel (a) presents the results from an ANOVA analysis. The full model includes the standardized major fixed effect ($b_{m,std}$) survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, and an indicator for parental college education, controls for occupation, graduate education, employment region, employer size, employer age, employer ownership structure, thirteen indicators for the type of primary and secondary job tasks (work activity) and an indicator for whether the worker supervises others. The full model sum of squares (SS) is 46121.55 and has a $R^2 = 0.49$. Each row corresponds to a separate regression which is the full model less the variables listed in the first column. The mode for *job level data* excludes by work tasks and supervisory work, the model *employer data* excludes employer size, ownership and age and the model *job or employer data* excludes both sets. Δ model SS (%)=$\frac{\text{model SS}|\text{row}-\text{model SS}|\text{all vars}}{\text{model SS}|\text{all vars}}$. R^2 is adjusted R-squared. Panel (b) presents F-statistics and P-values from the full log earnings model. These results from a test of joint significance for all variables in the group. Standard errors are clustered by major. Each regression uses weights equal to survey-weighted employment. Data source is the 2003-2019 NSCG. N= 250,265 in each regression.

Figure 3.4: Occupation Analysis

(a) Sorting of Majors across Occupations

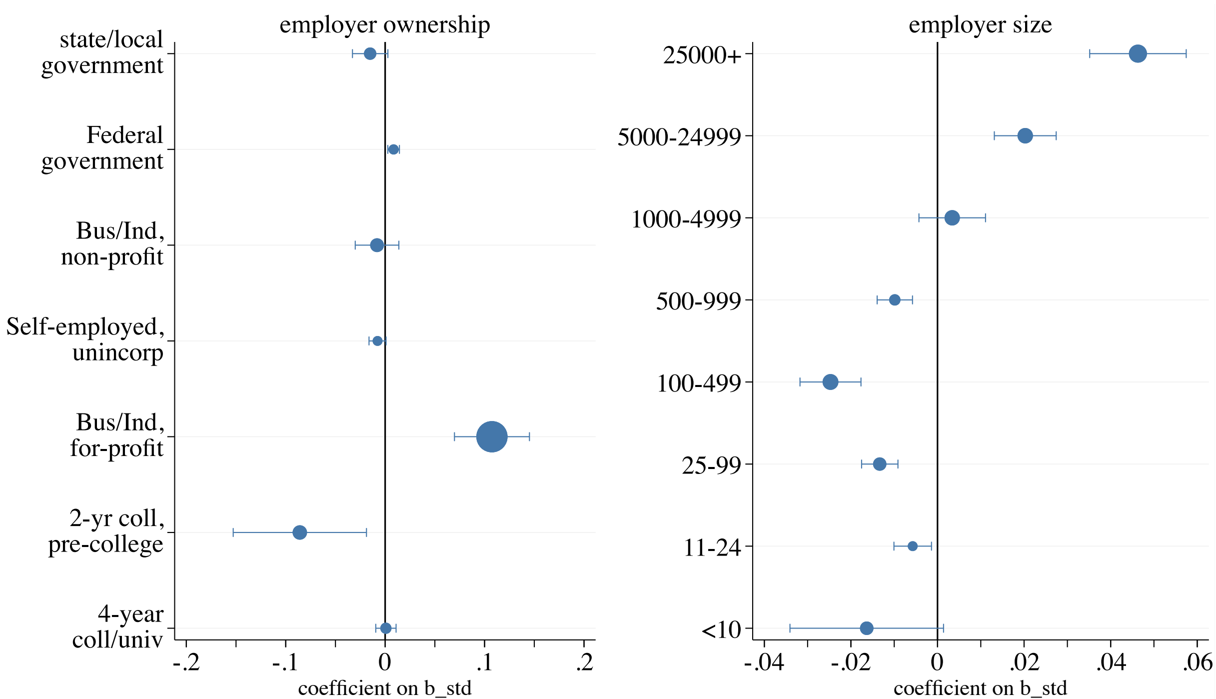


(b) Sorting and Occupation-Earnings



Note: Panel (a) illustrate that there is sorting of college majors across occupations. For each occupation, the figure plots the coefficient and 95% confidence interval for $b_{m,std}$ from a regressions with occupation dummies as the dependent variable and the baseline model covariates as the independent variables: $b_{m,std}$ survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, and an indicator for parental college education. The magnitude of the coefficient measures the percentage point difference in the probability of working in a particular occupation between the average major and majors that earn one standard deviation above the mean ($b_{m,std} = 1$). For each occupation, Panel (b) plots the coefficients in Panel (a), against the coefficients on the occupation fixed effects in the full model. The full model is a regression of log annual earnings the baseline covariates and adds in controls for occupation, graduate education, employment region, employer size, employer age, employer ownership structure, thirteen indicators for the type of primary and secondary job tasks (work activity) and an indicator for whether the worker supervises others. The size of the dot corresponds to the survey-weighted employment in the occupation. Data source is the 2003-2019 NSCG. N= 250,265 in each regression.

Figure 3.5: Employer Characteristics Analysis



Note: Figure illustrate that there is sorting of college majors across employer ownership structures and employer sizes. For each ownership categories, the figure plots the coefficient and 95% confidence interval for $b_{m,std}$ from a regressions with a dummy for the employer characteristic category as the dependent variable and the baseline model covariates as the independent variables: $b_{m,std}$ survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, and an indicator for parental college education. The magnitude of the coefficient measures the percentage point difference in the probability of working at that employer type between the average major and majors that earn one standard deviation above the mean ($b_{m,std} = 1$). The size of the dot corresponds to the survey-weighted employment in each category. Data source is the 2003-2019 NSCG. N= 250,265 in each regression.

Table 3.4: Robustness of the Gelbach Decomposition Results

	Main	No Cntrls	Broad Occ	No Wgt	Has PT Workers	Inst Type in Base	Inst Type in Full
Panel B: Gelbach Decomposition, Total Change Partitioned Across Variable Groups ($\hat{\delta}_g$ and 100 $\hat{\delta}_g/\hat{\delta}$)							
$\hat{\alpha}_1^{base}$	.1687***	.1963***	.1687***	.1620***	.1838***	.1559***	.1687***
	(.0011)	(.0000)	(.0011)	(.0013)	(.0014)	(.0047)	(.0011)
$\hat{\alpha}_1^{full}$	.0769***	.0771***	.0832***	.0772***	.0684***	.0761***	.0733***
	(.0042)	(.0041)	(.0045)	(.0035)	(.0044)	(.0043)	(.0044)
$\hat{\delta} : \hat{\alpha}_1^{base} - \hat{\alpha}_1^{full}$	.0918***	.1192***	.0854***	.0848***	.1153***	.0806***	.0953***
	(.0040)	(.0041)	(.0044)	(.0031)	(.0045)	(.0029)	(.0043)
% change in $\hat{\alpha}_1^{base}$	54%	61%	51%	52%	63%	51%	57%
Panel B: Gelbach Decomposition, Total Change Partitioned Across Variable Groups ($\hat{\delta}_g$ and 100 $\hat{\delta}_g/\hat{\delta}$)							
occupation	.0345***	.0386***	.0319***	.0314***	.0358***	.0306***	.0340***
	38%	32%	37%	37%	31%	38%	36%
grad educ	.0041	.0028	.0042	-.0064*	.0029	-.0082***	.0039
	5%	2%	5%	-8%	2%	-10%	4%
emp region	.0006	.0015	.0006	.0006	.0003	.0004	.0005
	1%	1%	1%	1%	0%	0%	1%
emp age	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	0%	0%	0%	0%	0%	0%	0%
emp ownership	.0242***	.0318***	.0222***	.0324***	.0261***	.0321***	.0238***
	26%	27%	26%	38%	23%	40%	25%
emp size	.0122***	.0083***	.0116***	.0139***	.0111***	.0138***	.0120***
	13%	7%	14%	16%	10%	17%	13%
job level	.0042***	.0061***	.0043***	.0010	.0061***	.0006	.0042***
	5%	5%	5%	1%	5%	1%	4%
primary tasks	.0098***	.0079**	.0085***	.0087***	.0123***	.0085***	.0097***
	11%	7%	10%	10%	11%	11%	10%
secondary tasks	.0023***	.0029***	.0021***	.0031***	.0030***	.0029***	.0022***
	2%	2%	2%	4%	3%	4%	2%
weeks/year: 41+	–	–	–	–	.0071***	–	–
	–	–	–	–	6%	–	–
hours/week: 34+	–	–	–	–	.0107**	–	–
		–	–	–	9%	–	–
BA institution	–	–	–	–	–	–	.0051***
		–	–	–	–	–	5%
baseline vars	–	.0193***	–	–	–	–	–
		16%	–	–	–	–	

Note: Table presents the results of the Gelbach (2016) decomposition. Each column corresponds to a different specification. Panel (a) presents the coefficient on $b_{m,std}$ in the baseline and full model, and the total change in the coefficient between the models ($\hat{\delta}$). Panel (b) presents, for each covariate group, the log points accounted for by the covariate group g conditional on all of the other covariates simultaneously ($\hat{\delta}_g$) and percent of the total change accounted for by the covariate group g ($100 \times \hat{\delta}_g/\hat{\delta}$). Column 1 is the primary specification. Column 2 constructs $\hat{b}_{m,std}$ from a baseline model with no controls (only the major fixed effects). Column 3 uses 20 broad occupation codes in the full model instead of 80 detailed occupation codes Column 4 is the main specification without sample weights. Column 5 adds in employed workers that work less than 35 hours per week (full-time) and 41 weeks per year (full-year), and controls for this in the full model. Column 6 adds in controls for BA institution type in the baseline model used to construct $\hat{b}_{m,std}$ and Column 7 instead just adds in controls for BA institution type into the full model. Standard errors are clustered by major. Decomposition performed using the stata command *b1x2*. All regressions (expect Column 4) uses weights equal to survey-weighted employment. Data source is the 2003-2019 NSCG. N= 250,265 in each regression except in Column 4 where N=300,506.

APPENDICES

APPENDIX A

Appendix to Chapter 1

A.1 Additional Tables & Figures Appendix

Table A.1: Person-Year Summary Statistics for the Main Analysis Sample

	ACS-LEHD				NSCG-LEHD (weighted)			
college major group:	all	general	not general not specific	specific	all	general	not general not specific	specific
employers								
mean employers in year	1.54	1.57	1.53	1.52	1.52	1.56	1.45	1.49
% with only one employer in year	64%	62%	64%	65%	65%	64%	68%	66%
mean employers in year w/ earn > min wage thres	0.97	0.96	0.96	0.99	1.01	1.00	0.98	1.05
% with only one in year w/ earn > min wage thres	81%	79%	80%	82%	83%	82%	82%	84%
employer change	26%	28%	27%	22%	23%	25%	23%	19%
2 digit industry change	16%	18%	18%	13%	14%	17%	14%	10%
4 digit industry change	20%	22%	21%	16%	17%	20%	17%	13%
total annual earnings								
mean	48,850	47,060	47,700	52,340	52,370	50,420	49,430	57,950
mean log	10.53	10.48	10.50	10.62	10.64	10.59	10.57	10.78
percent earned at main job	91%	91%	91%	92%	92%	91%	93%	93%
% w/ main job earnings > 90% of total earnings	75%	73%	75%	77%	77%	74%	79%	77%
% w/ earnings in multiple states in year	3.8%	4.1%	3.9%	3.4%	3.7%	4.6%	2.9%	3.0%
duration between an individual's observations								
mean	1.11	1.12	1.12	1.09	1.08	1.09	1.10	1.06
percent duration = 1 year	95%	95%	95%	96%	96%	95%	97%	97%
observations	2,637,000	1,172,000	648,000	817,000	76,000	37,000	17,000	22,000
individuals	383,000	174,000	112,000	96,500	11,500	4,400	3,000	4,100

Note: Table displays summary statistics for all person-year observations in the main ACS-LEHD and NSCG-LEHD analysis samples. Annual observations are only included in the worker had at least 3+ quarters of non-zero earnings in the year in one of the 23 covered states (see Appendix Figure A.10). Total annual earnings defined as earnings summed across all jobs in a year. Main job is defined as the employer at which the worked had the highest annual earnings in the year. Minimum wage threshold is defined as the prevailing minimum wage x 35 hours x 40 weeks. NSCG-LEHD summary statistics are weighted using NSCG survey weights. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. "D" represents cells that have been deleted during disclosure review.

Table A.2: College Major Specificity Measures

Panel A: Specific Majors

College Major	% in top 3 occs	% in top 5 occs	Herfindahl Index
Nursing	0.92	0.93	0.794
Special Education and Teaching	0.77	0.82	0.296
Teacher Education	0.70	0.72	0.267
Accounting	0.69	0.71	0.413
Other Education	0.68	0.72	0.312
Library Science	0.66	0.76	0.263
Pharmacy	0.65	0.70	0.324
Civil Engineering	0.61	0.68	0.200
Computer Engineering	0.58	0.64	0.212
Social Work	0.51	0.54	0.175
Architecture	0.48	0.56	0.108
Electrical, Electronics Engineering	0.48	0.55	0.085
Computer and Information Sciences	0.47	0.56	0.117
Mechanical engineering	0.44	0.54	0.090
Aeronautical Engineering	0.44	0.56	0.097
Chemical engineering	0.40	0.48	0.071
Nutritional sciences	0.38	0.42	0.096
Rehabilitation and Therapeutic Professions	0.38	0.47	0.059
Allied Health	0.35	0.40	0.086
Materials Science and Engineering	0.35	0.45	0.054

Panel B: Not General or Specific Majors

College Major	% in top 3 occs	% in top 5 occs	Herfindahl Index
Applied Arts	0.34	0.40	0.078
Atmospheric Sciences and Meteorology	0.34	0.41	0.087
Legal Studies	0.33	0.39	0.064
Industrial, Manufacturing Engineering	0.32	0.41	0.054
Geology and Earth Science	0.31	0.36	0.072
Statistics	0.31	0.43	0.051
Finance	0.30	0.38	0.047
Mathematics	0.27	0.35	0.036
Chemistry	0.26	0.35	0.045
Protective Services	0.25	0.32	0.036

Note: Table displays the percent of each major's total employment accounted for by the three largest occupations for the major and the Herfindahl index based on the occupation employment shares: $H_m \sum_o (E_{mo})^2$ where E_{mo} is the share of major m graduates employed in occupation o. Data source is the public-use 2009-2019 ACS. Employed workers are restricted to unenrolled college graduates who are employed 1-5 years post undergraduate degree and that work at least part-time part-year. The 20 majors with graduates that are clustered in a small number of occupations (i.e. a high percent of graduates employed in the major's three largest occupations) are considered specific majors. General majors are the 20 majors with the widest dispersion of graduates across occupations.

Continued: College Major Specificity Measures

Panel B (Continued): Not General or Specific Majors

College Major	% in top 3 occs	% in top 5 occs	Herfindahl Index
Other Engineering	0.25	0.34	0.034
Family and Consumer Sciences	0.25	0.34	0.033
Biomedical Engineering	0.24	0.34	0.035
Biochemistry, Biophysics, Molecular Biology	0.24	0.33	0.032
Microbiology	0.24	0.36	0.035
Management Information Systems and Science	0.23	0.35	0.036
Philosophy, Religion & Theology	0.22	0.27	0.032
Agriculture	0.20	0.26	0.024
Marketing	0.20	0.29	0.027
Psychology	0.20	0.26	0.022

Panel C: General Majors

College Major	% in top 3 occs	% in top 5 occs	Herfindahl Index
Journalism	0.19	0.27	0.026
Physics	0.19	0.27	0.025
Public Policy	0.18	0.26	0.023
Economics	0.18	0.25	0.023
Health and Medical Administrative Services	0.17	0.24	0.021
Engineering technology	0.17	0.25	0.021
Foreign Language & Linguistics	0.17	0.23	0.020
Fitness, Recreation and Leisure Studies	0.16	0.23	0.019
English, Liberal Arts, Humanities	0.16	0.22	0.019
Sociology	0.16	0.22	0.018
Geography	0.15	0.20	0.018
Biology	0.15	0.21	0.017
Political Science, Government, Int'l Relations	0.14	0.20	0.017
Business, general	0.14	0.21	0.018
Other Physical Sciences	0.13	0.17	0.013
Public Administration	0.13	0.19	0.018
Natural Resources	0.13	0.18	0.013
Communications	0.12	0.18	0.015
Public Health	0.12	0.18	0.016
Other Visual/Performing Arts	0.12	0.19	0.016
Other Social Sciences	0.10	0.16	0.013

Note: Table displays the percent of each major's total employment accounted for by the three largest occupations for the major and the Herfindahl index based on the occupation employment shares: $H_m \sum_o (E_{mo})^2$ where E_{mo} is the share of major m graduates employed in occupation o. Data source is the public-use 2009-2019 ACS. Employed workers are restricted to unenrolled college graduates who are employed 1-5 years post undergraduate degree and that work at least part-time part-year. The 20 majors with graduates that are clustered in a small number of occupations (i.e. a high percent of graduates employed in the major's three largest occupations) are considered specific majors. General majors are the 20 majors with the widest dispersion of graduates across occupations.

Table A.3: Top Three Most Common Occupations for Select Majors

General Majors		**Not General, Not Specific**		**Specific Majors**	
Other Visual/Performing Arts		*Microbiology*		*Nursing*	
Waiters & Waitresses	.045	Clinical Laboratory Technologists	.089	Registered Nurses	.891
Other Teachers & Instructors	.039	Physicians & Surgeons	.079	Nurse Practitioners & Nurse Midwives	.016
Elementary & Middle School Teachers	.037	Biological Scientists	.069	Home Health Aides	.014
Communications		*Protective Services*		*Accounting*	
Marketing & Sales Managers	.045	Police Officers	.149	Accountants & Auditors	.641
Customer Service Representatives	.044	Security Guards & Gaming Surveillance Officers	.056	Financial Managers	.024
Secretaries & Administrative Assistants	.034	Bailiffs, Correctional Officers, & Jailers	.047	Bookkeeping & Auditing Clerks	.023
Natural Resources		*Mathematics*		*Chemical engineering*	
Environmental Scientists & Geoscientists	.048	Elementary & Middle School Teachers	.109	Chemical Engineers	.206
Conservation Scientists & Foresters	.040	Secondary School Teachers	.097	Miscellaneous Engineers	.101
Miscellaneous Managers	.039	Software Developers	.059	Industrial Engineers	.091
Business		*Applied Arts*		*Computer & Information Sciences*	
Accountants & Auditors	.063	Designers	.260	Software Developers	.307
Supervisors of Retail Sales Workers	.043	Retail Salespersons	.045	Computer Programmers	.104
Miscellaneous Managers	.037	Supervisors of Retail Sales Workers	.035	Computer Support Specialists	.057
Sociology		*Geology & Earth Science*		*Teacher Education*	
Social Workers	.078	Environmental Scientists & Geoscientists	.253	Elementary &Middle School Teachers	.470
Counselors	.043	Miscellaneous Managers	.032	Secondary School Teachers	.209
Elementary & Middle School Teachers	.038	Retail Salespersons	.026	Preschool & Kindergarten Teachers	.018
Other Social Sciences		*Finance*		*Social Work*	
Elementary & Middle School Teachers	.036	Accountants & Auditors	.170	Social Workers	.407
Social Workers	.033	Financial Managers	.066	Counselors	.067
Secretaries & Administrative Assistants	.032	Personal Financial Advisors	.059	Other Therapists	.034
Poli Science, Gov & Int'l Relations		*Family & Consumer Sciences*		*Nutritional sciences*	
Lawyers, & judges, judicial workers	.072	Elementary & Middle School Teachers	.114	Dieticians & Nutritionists	.295
Miscellaneous Managers	.041	Preschool & Kindergarten Teachers	.077	Registered Nurses	.052
Paralegals & Legal Assistants	.029	Childcare Workers	.055	Recreation & Fitness Workers	.032

Note: For each major, table displays the occupation and percent of each major's total employment accounted for by the three largest occupations for the major. Data source is the public-use 2009-2019 ACS. Each columns correspond to a college major specificity group, categorized by the percent of the major's graduates employed in the top three occupations. Employed individuals include unenrolled college graduates in the first five years of the career from the 1999-2012 graduating cohorts who are employed at least part-time part-year.

Table A.4: Characteristics of General and Specific Majors

	general majors	not general not specific	specific majors	all majors
Panel A: Major Subject Field (N majors)				
Business & Economics	2	2	1	5
Communications & Marketing	2	1	0	3
Computer Science & Engineering	0	2	9	11
Education	0	0	3	3
Health	2	0	4	6
Humanities	3	2	0	5
Pure Sciences	4	9	1	14
Social Sciences	6	2	1	9
All Other Majors	2	2	1	5
Panel B: Institution Type (%)				
Research University I (RI)	26	25	22	25
Liberal Arts	14	11	9	12
Public	65	69	71	68
Panel C: Parent's Education (%)				
1+ parent has BA	63	60	60	61
Panel D: Self-Report Match between Major and Job (%)				
closely related	30	47	74	47
somewhat	38	26	17	29
not related	32	26	9	24

Note: Panel A displays the distribution of college major subject fields within general and specific major. Each cell contains the number of college majors. Columns correspond to the three mutually exclusive categories of college major specificity. The three groups of college major specificity are defined using the occupational dispersion of recent college graduates (see Figure 1.1). Rows correspond to 9 mutually exclusive subject field groupings of majors. See Appendix A.16 for full crosswalk of the 61 college majors to the 9 subject fields. Panel B reports the percent of individuals that attended each institution type according to the 1994 Carnegie Classification. Panel C reports the percent of individuals with at least one parent with a BA degree. Panel D reports the percent of individuals who report that their primary job is closely, somewhat or not related to their primary job. Estimates in Panel B-D come from the public-use 2010-2019 NSCG. Sample is restricted to new respondents to each NSCG wave from the graduating cohorts of 1999-2015. Panel D further restricts sample to individuals who are less than 5 years post graduation and don't have a graduate degree.

Table A.5: Total Years in Analysis Sample

cohort	general major	not general or specific	specific major
1999-2000	10.50	10.50	11.26
2001-2002	9.25	9.18	9.83
2003-2004	7.90	7.94	8.52
2005-2006	6.57	6.62	7.08
2007-2008	5.20	5.24	5.56
2009-2010	3.77	3.78	3.98
2011-2012	2.27	2.26	2.35

Note: Table provides average number of years in the analysis sample by graduation cohort and college major specificity group. Cohort is two-year bins of bachelor's degree graduation year. For each worker, the total years in the analysis sample is equal to the total years a worker had three-plus quarters of non-zero LEHD earnings in the 23 covered states through year 2014. The total possible years in the analysis sample spans from 2 for the 2012 cohort to 15 years for the 1999 cohort. The dataset is the ACS-LEHD. The sample includes all individuals in the main analysis sample. Results were disclosed by the U.S Census Bureau's Disclosure Review Board with approval number CBDRB-FY22-P2420-R9606. All point estimates are rounded.

Table A.6: Composition of General and Specific Majors, by Gender

	female			male		
	general	not general or specific	specific	general	not general or specific	specific
Computer Science/Engineering	-	2.99	14.2	-	10.3	65.7
Business & Economics	24.8	7.38	9.9	33.9	17.3	10.5
Communications/Marketing	13.9	11.2	-	10.2	9.40	-
Education	-	-	41.9	-	-	15.8
Health	1.90	-	27.3	0.65	-	6.68
Humanities	23.8	17.6	-	19.6	17.1	-
Other	3.64	7.82	0.11	6.72	10.9	0.03
Pure Sciences	15.3	16.2	1.13	14.6	22.5	0.26
Social Sciences	16.7	36.9	5.47	14.3	12.6	0.96

Note: Table displays that the distribution of general (specific) majors across major subject field separately for males and females. Each row corresponds to a major subject field and each column corresponds to a college major specificity category. College major subject fields are grouped into nine mutually exclusive college major subject fields (see Appendix Table A.16). College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). Data source is the public-use ACS, 2009-2019. Entry in each cell is the (weighted) percent of column with the row's major subject field.

Table A.7: Regression Coefficients: Log Annual Earnings

	Log Annual Earnings						
years post grad	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general		0.2658***	0.4096***	0.5089***	0.5764***	0.6175***	0.6586***
		(0.0019)	(0.0031)	(0.0044)	(0.0057)	(0.0072)	(0.0096)
not gen/spec	0.0373***	-0.0114***	-0.0177***	-0.0220***	-0.0185***	-0.0245***	-0.0375***
	(0.0026)	(0.0025)	(0.0032)	(0.0038)	(0.0046)	(0.0058)	(0.0080)
specific	0.1817***	-0.0031	-0.0378***	-0.0624***	-0.0835***	-0.1011***	-0.1188***
	(0.0026)	(0.0023)	(0.0030)	(0.0035)	(0.0041)	(0.0051)	(0.0070)
Panel B: earnings gap: (major m - general)							
not gen/spec	.0373***	.0259***	.0195***	.0152***	.0188***	.0127**	-.0003
	(.0026)	(.0028)	(.0031)	(.0036)	(.0043)	(.0056)	(.0079)
specific	.1817***	.1786***	.1439***	.1193***	.0982***	.0806***	.0628***
	(.0026)	(.0026)	(.0028)	(.0032)	(.0039)	(.0049)	(.0068)
Panel C: earnings growth from previous period							
general		.2658***	.1438***	.0993***	.0675***	.0412***	.0410***
		(.0019)	(.0020)	(.0022)	(.0024)	(.0029)	(.0045)
not gen/not spec		.2544***	.1375***	.0950***	.0711***	.0351***	.0280***
		(.0023)	(.0025)	(.0026)	(.0030)	(.0037)	(.0057)
specific		.2627***	.1090***	.0747***	.0464***	.0236***	.0233***
		(.0021)	(.0022)	(.0023)	(.0025)	(.0031)	(.0047)
Panel D: earnings growth gap: (major m - general)							
not gen/spec		-.0114***	-.0063**	-.0043	.0036	-.0061	-.0130**
		(.0025)	(.0026)	(.0029)	(.0034)	(.0043)	(.0064)
specific		-.0031	-.0348***	-.0246***	-.0211***	-.0176***	-.0177***
		(.0023)	(.0023)	(.0025)	(.0030)	(.0037)	(.0055)

Note: Table displays estimates from a regression of log annual earnings on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College majors are defined as either general, specific and not general or specific majors (not gen/spec). See Figure 1.1). All regression include controls as described in Section 1.3. In each year since graduation bin g, Panel A provides the estimates and standard errors of β_g for general majors and $\beta_{m,g}$ for other majors m. Panel B-D provide coefficients and standard errors on linear combinations of estimates in Panel A. In Panel B the earnings gap between major m and general majors is calculated as $\beta_{m,g} + \phi_m$. Panel C displays earnings growth from period $g-1$ to g which is $(\beta_g - \beta_{g-1})$ for general majors and $(\beta_{m,g} - \beta_{m,g-1}) + (\beta_g - \beta_{g-1})$ for major m. Panel D displays the difference in earnings growth between major m and general majors, which is $\beta_{m,g} - \beta_{m,g-1}$. Standard errors are clustered at the individual level. Data source is the ACS-LEHD, see Section 1.2.2 for details on the analysis sample. Regression includes 2,637,000 observations for 383,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Table A.8: Regression Coefficients: Annual Earnings

	Annual Earnings ($)						
years post grad	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general	30270	8730***	14440***	20250***	25330***	30380***	37380***
		(570.8)	(589.3)	(831.1)	(1079)	(1355)	(1563)
not gen/spec	1523***	-911.3	-111.2	28.22	335.2	-83.62	-1818**
	(124.5)	(751.3)	(319.8)	(436)	(477.7)	(539.5)	(764.7)
specific	7501***	-139.9	-349.3	-963.2***	-2427***	-3926***	-5297***
	(148.2)	(762)	(230.9)	(324.5)	(315.5)	(423.4)	(798.8)
Panel B: earnings gap: (major m - general)							
not gen/spec	1523***	611.4	1411***	1551***	1858***	1439**	-295
	(124.5)	(795.2)	(393.2)	(507.8)	(542.2)	(588)	(761.7)
specific	7501***	7362***	7152***	6538***	5074***	3576***	2204***
	(148.2)	(753.1)	(241.2)	(332.7)	(324.5)	(427)	(809.7)
Panel C: earnings growth from previous period							
general		8730***	5714***	5808***	5077***	5048***	7005***
		(570.8)	(1092)	(281.9)	(293)	(334.8)	(484.9)
not gen/spec		7819***	6514***	5948***	5384***	4630***	5271***
		(216.6)	(380.7)	(292.3)	(298.3)	(396.8)	(598.4)
specific		8590***	5505***	5194***	3613***	3550***	5633***
		(278.3)	(328.8)	(363.6)	(335.4)	(385.2)	(622.1)
Panel D: earnings growth gap: (major m - general)							
not gen/spec		-911.3	800.1	139.5	306.9	-418.8	-1734**
		(751.3)	(835.2)	(217.4)	(222.3)	(337.2)	(711.3)
specific		-139.9	-209.4	-613.9**	-1464***	-1499***	-1372**
		(762)	(811)	(287.4)	(275.4)	(326.1)	(646.5)

Note: Table displays estimates from a regression of annual earnings on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College majors are defined as either general, specific and not general or specific majors (not gen/spec). See Figure 1.1). All regression include controls as described in Section 1.3. In each year since graduation bin g, Panel A provides the estimates and standard errors of β_g for general majors and $\beta_{m,g}$ for other majors m. Panel B-D provide coefficients and standard errors on linear combinations of estimates in Panel A. In Panel B the earnings gap between major m and general majors is calculated as $\beta_{m,g} + \phi_m$. Panel C displays earnings growth from period $g-1$ to g which is $(\beta_g - \beta_{g-1})$ for general majors and $(\beta_{m,g} - \beta_{m,g-1}) + (\beta_g - \beta_{g-1})$ for major m. Panel D displays the difference in earnings growth between major m and general majors, which is $\beta_{m,g} - \beta_{m,g-1}$. Standard errors are clustered at the individual level. Data source is the ACS-LEHD, see Section 1.2.2 for details on the analysis sample. Regression includes 2,637,000 observations for 383,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Table A.9: Regression Coefficients: Employer and Industry Changing

Employer Change

years post grad:	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general	.3284	-0.0134***	-0.0459***	-0.0728***	-0.0944***	-0.1117***	-0.1322***
		(0.0016)	(0.0018)	(0.0022)	(0.0026)	(0.0032)	(0.0041)
not gen/spec	-0.0057***	-0.0026	-0.0039	-0.0053*	-0.0062**	-0.0061*	0.0010
	(0.0021)	(0.0025)	(0.0026)	(0.0027)	(0.0029)	(0.0031)	(0.0036)
specific	-0.0486***	-0.0098***	-0.0090***	-0.0035	0.0027	0.0086***	0.0176***
	(0.0019)	(0.0023)	(0.0024)	(0.0024)	(0.0026)	(0.0028)	(0.0032)
Panel B: implied job changing rate							
general	0.3474	0.3340	0.3015	0.2746	0.2530	0.2357	0.2152
not gen/spec	0.3417	0.3257	0.2919	0.2636	0.2411	0.2239	0.2105
specific	0.2988	0.2755	0.2439	0.2225	0.2071	0.1957	0.1842
Panel C: job changing gap: (major m - general)							
not gen/spec	-.0057***	-.0083***	-.0096***	-.011***	-.0119***	-.0118***	-.0047
	(.0021)	(.0016)	(.0017)	(.0018)	(.002)	(.0024)	(.003)
specific	-.0486***	-.0584***	-.0576***	-.0521***	-.0459***	-.04***	-.031***
	(.0019)	(.0015)	(.0015)	(.0016)	(.0018)	(.0021)	(.0027)

Two-Digit Industry Change

years post grad:	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general	.2512	-0.0194***	-0.0491***	-0.0720***	-0.0897***	-0.1014***	-0.1150***
		(0.0014)	(0.0016)	(0.0019)	(0.0022)	(0.0027)	(0.0035)
not gen/spec	-0.0077***	-0.0003	-0.0004	-0.0011	-0.0003	0.0015	0.0043
	(0.0019)	(0.0023)	(0.0023)	(0.0024)	(0.0025)	(0.0027)	(0.0031)
specific	-0.0598***	-0.0043**	0.0038*	0.0122***	0.0203***	0.0259***	0.0307***
	(0.0017)	(0.0020)	(0.0020)	(0.0021)	(0.0022)	(0.0023)	(0.0026)
Panel B: implied job changing rate							
general	0.2512	0.2318	0.2021	0.1792	0.1615	0.1498	0.1362
not gen/spec	0.2435	0.2238	0.1940	0.1704	0.1535	0.1436	0.1328
specific	0.1914	0.1677	0.1461	0.1316	0.1220	0.1160	0.1072
Panel C: job changing gap: (major m - general)							
not gen/spec	-.0077***	-.0081***	-.0082***	-.0088***	-.0081***	-.0062***	-.0034
	(.0019)	(.0014)	(.0015)	(.0015)	(.0017)	(.002)	(.0025)
specific	-.0598***	-.0641***	-.056***	-.0476***	-.0395***	-.0338***	-.029***
	(.0017)	(.0013)	(.0013)	(.0013)	(.0014)	(.0017)	(.0021)

Note: Table displays estimates from a regression of employer or industry change on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College majors are defined as either general, specific and not general or specific majors (not gen/spec). See Figure 1.1). All regression include controls as described in Section 1.3. In each year since graduation bin g, Panel A provides the estimates and standard errors of β_g for general majors and $\beta_{m,g}$ for other majors m. Panel B-C provide coefficients and standard errors on linear combinations of estimates in Panel A. In Panel B the job changing rate is calculated as $\beta_0 + \beta_g$ for general majors and as $\beta_0 + \beta_m + \beta_g + \beta_{m,g}$ for specific majors, where β_0 is the sample mean of the outcome among general majors in the omitted period. In Panel B the job changing gap between major m and general majors is calculated as $\beta_{m,g} + \phi_m$. Standard errors are clustered at the individual level. Data source is the ACS-LEHD, see Section 1.2.2 for details on the analysis sample. Regression includes 2,254,000 observations for 360,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Table A.10: Regression Coefficients: (Unweighted) Occupation Changing

Narrow Occupation Change (Unweighted)						
years post grad:	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:						
general	.5862	-0.0329***	-0.0778***	-0.1178***	-0.1307***	-0.1577***
		(0.0119)	(0.0139)	(0.0155)	(0.0191)	(0.0224)
not gen/spec	-0.0259*	-0.0256	-0.0070	-0.0205	0.0001	-0.0077
	(0.0138)	(0.0176)	(0.0176)	(0.0181)	(0.0188)	(0.0191)
	-0.1609***	-0.0207	0.0330**	0.0814***	0.0876***	0.1116***
specific	(0.0132)	(0.0163)	(0.0165)	(0.0168)	(0.0175)	(0.0176)
Panel B: implied job changing rate						
general	0.5862	0.5533	0.5084	0.4684	0.4555	0.4285
not gen/spec	0.5603	0.5019	0.4755	0.4220	0.4298	0.3950
specific	0.4253	0.3717	0.3805	0.3889	0.3822	0.3792
Panel C: job changing gap: (major m - general)						
not gen/spec	-.0259*	-.0514***	-.0328***	-.0464***	-.0257**	-.0335**
	(.0138)	(.0119)	(.0112)	(.0119)	(.0128)	(.0132)
specific	-.1609***	-.1815***	-.1279***	-.0795***	-.0733***	-.0493***
	(.0132)	(.0109)	(.0103)	(.0106)	(.0116)	(.0118)
Broad Occupation Change (Unweighted)						
years post grad :	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:						
	3-4	5-6	7-8	9-10	11-12	13-15
general	.5225	-0.0410***	-0.0870***	-0.1307***	-0.1436***	-0.1601***
		(0.0126)	(0.0145)	(0.0160)	(0.0192)	(0.0224)
not gen/spec	-0.0450***	-0.0272	-0.0022	-0.0012	0.0103	0.0130
	(0.0149)	(0.0185)	(0.0186)	(0.0189)	(0.0195)	(0.0197)
specific	-0.2445***	0.0054	0.0590***	0.1232***	0.1187***	0.1547***
	(0.0137)	(0.0165)	(0.0168)	(0.0170)	(0.0176)	(0.0177)
Panel B: implied job changing rate						
general	0.5225	0.4815	0.4355	0.3918	0.3789	0.3624
not gen/spec	0.4775	0.4092	0.3882	0.3456	0.3442	0.3304
specific	0.2780	0.2424	0.2499	0.2705	0.2531	0.2726
Panel C: job changing gap: (major m - general)						
not gen/spec	-.045***	-.0723***	-.0473***	-.0462***	-.0347***	-.032**
	(.0149)	(.0123)	(.0115)	(.0118)	(.0126)	(.0129)
specific	-.2445***	-.2392***	-.1855***	-.1214***	-.1258***	-.0898***
	(.0137)	(.0109)	(.0101)	(.0104)	(.0112)	(.0114)

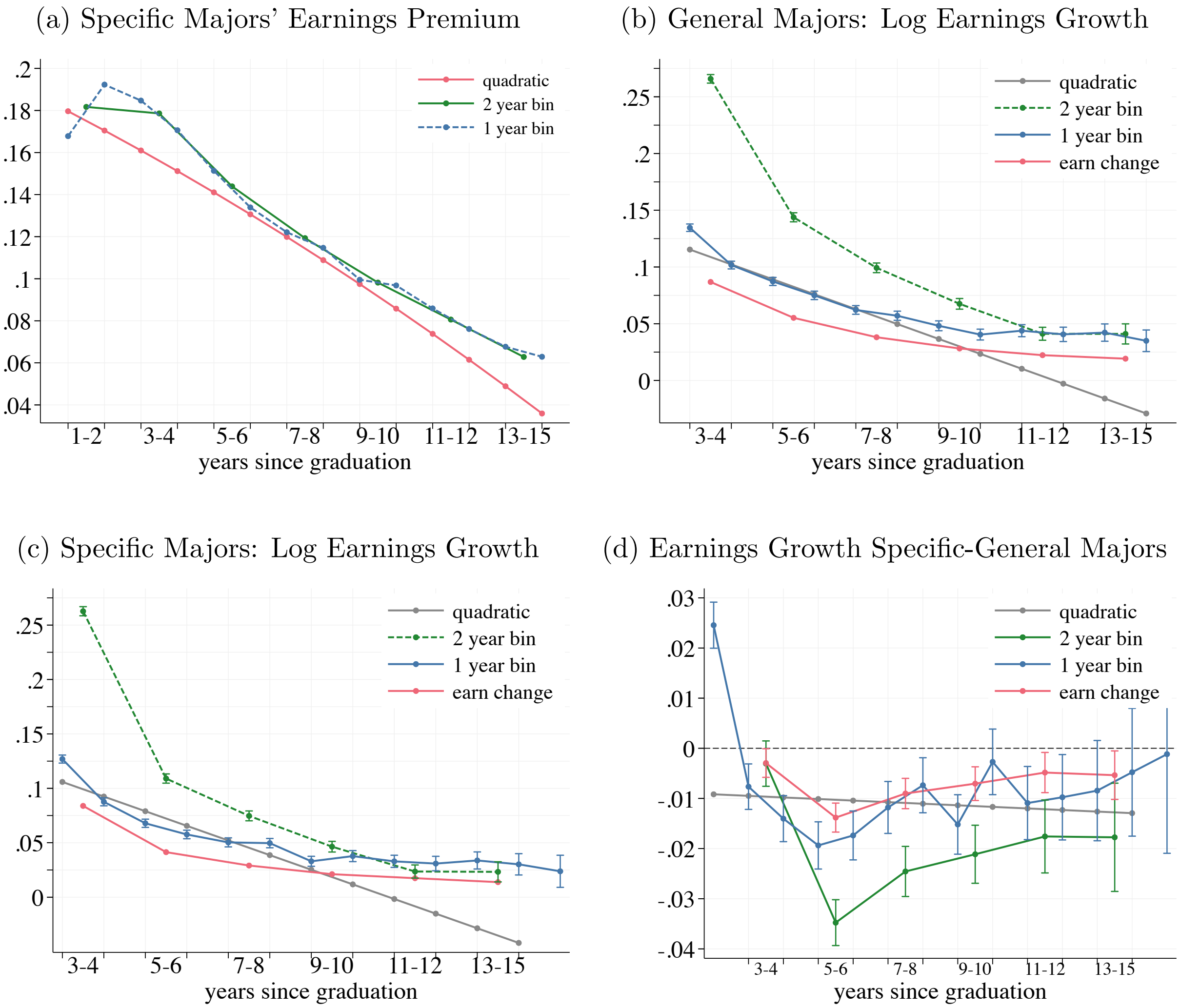
Note: Table displays estimates from a regression of narrow and broad occupation change on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). There are 20 broad and 80 narrow categories (see Appendix Section A.2.3 for definitions). All regression include controls as described in Section 1.3. In each year since graduation bin g, Panel A provides the estimates and standard errors of β_g for general majors and $\beta_{m,g}$ for other majors m. Panel B-C provide coefficients and standard errors on linear combinations of estimates in Panel A. In Panel B the job changing rate is calculated as $\beta_0 + \beta_g$ for general majors and as $\beta_0 + \beta_m + \beta_g + \beta_{m,g}$ for specific majors, where β_0 is the sample mean of the outcome among general majors in the omitted period. In Panel B the job changing gap between major m and general majors is calculated as $\beta_{m,g} + \phi_m$. Standard errors are clustered at the individual level. Data source is the ACS-NSCG. Regression includes 61,500 observations for 35,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Table A.11: Regression Coefficients: (Weighted) Occupation Changing

Narrow Occupation Change (Weighted)						
years post grad :	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:						
general	0.7051	-0.0172	-0.1154***	-0.1553***	-0.1590***	-0.1494**
		(0.0369)	(0.0418)	(0.0442)	(0.0533)	(0.0623)
not gen/spec	0.0301	-0.1438**	-0.0376	-0.0466	-0.0516	-0.0671
	(0.0411)	(0.0560)	(0.0542)	(0.0536)	(0.0549)	(0.0545)
specific	-0.1976***	-0.0127	0.0594	0.1418***	0.0936*	0.1068**
	(0.0409)	(0.0518)	(0.0530)	(0.0505)	(0.0509)	(0.0499)
Panel B: implied job changing rate						
general	0.7051	0.6879	0.5897	0.5498	0.5461	0.5557
not gen/spec	0.7352	0.5741	0.5822	0.5333	0.5246	0.5186
specific	0.5075	0.4776	0.4515	0.4940	0.4421	0.4649
Panel C: job changing gap: (major m - general)						
not gen/spec	.0301	-.1137***	-.0075	-.0165	-.0215	-.0371
	(.0411)	(.0399)	(.0357)	(.0352)	(.0366)	(.036)
specific	-.1976***	-.2103***	-.1382***	-.0558*	-.104***	-.0908***
	(.0409)	(.0343)	(.0316)	(.0305)	(.0312)	(.0292)
Broad Occupation Change (Weighted)						
years post grad :	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:						
	3-4	5-6	7-8	9-10	11-12	13-15
general	0.5669	0.0034	-0.0827*	-0.1281***	-0.1713***	-0.1536**
		(0.0385)	(0.0429)	(0.0452)	(0.0540)	(0.0635)
not gen/spec	0.0598	-0.1735***	-0.0914	-0.0954*	-0.0745	-0.0904
	(0.0464)	(0.0578)	(0.0576)	(0.0574)	(0.0583)	(0.0580)
specific	-0.1855***	-0.0254	0.0177	0.0734	0.0697	0.0549
	(0.0413)	(0.0506)	(0.0511)	(0.0497)	(0.0500)	(0.0491)
Panel B: implied job changing rate						
general	0.5669	0.5703	0.4842	0.4388	0.3956	0.4133
not gen/spec	0.6267	0.4566	0.4526	0.4032	0.3809	0.3827
specific	0.3814	0.3593	0.3164	0.3267	0.2798	0.2827
Panel C: job changing gap: (major m - general)						
not gen/spec	.0598	-.1137***	-.0316	-.0356	-.0147	-.0306
	(.0464)	(.0403)	(.036)	(.034)	(.0355)	(.035)
specific	-.1855***	-.2109***	-.1678***	-.1121***	-.1158***	-.1306***
	(.0413)	(.0342)	(.0302)	(.029)	(.0294)	(.0272)

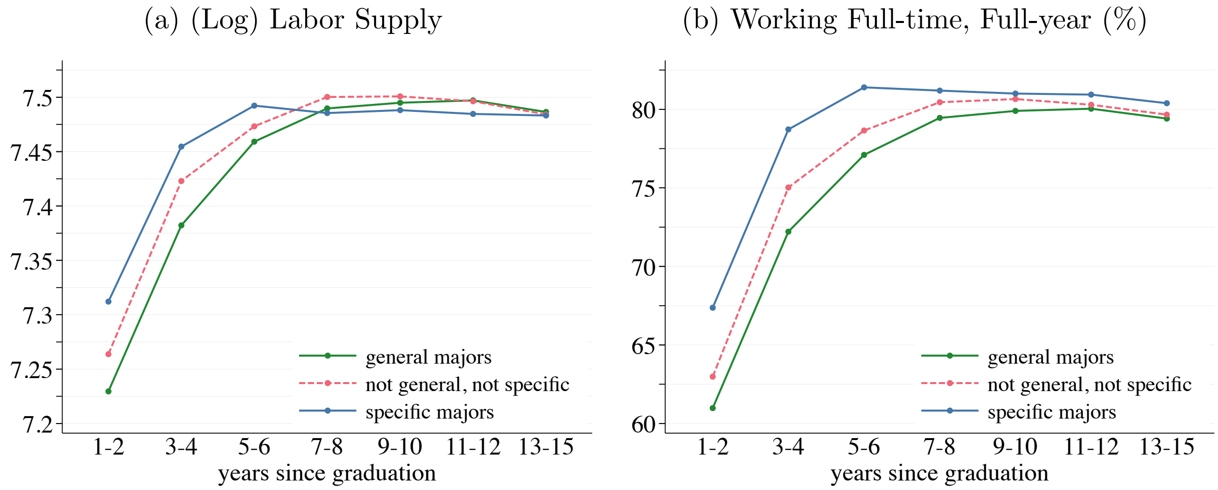
Note: Table displays estimates from a regression of narrow and broad occupation change on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). There are 20 broad and 80 narrow categories (see Appendix Section A.2.3 for definitions). All regression include controls as described in Section 1.3. In each year since graduation bin g, Panel A provides the estimates and standard errors of β_g for general majors and $\beta_{m,g}$ for other majors m. Panel B-C provide coefficients and standard errors on linear combinations of estimates in Panel A. In Panel B the job changing rate is calculated as $\beta_0 + \beta_g$ for general majors and as $\beta_0 + \beta_m + \beta_g + \beta_{m,g}$ for specific majors, where β_0 is the sample mean of the outcome among general majors in the omitted period. In Panel B the job changing gap between major m and general majors is calculated as $\beta_{m,g} + \phi_m$. Standard errors are clustered at the individual level. Data source is the ACS-NSCG. Regression includes 61,500 observations for 35,000 individuals. NSCG survey weights used. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Figure A.1: Robustness of Earnings Results to Model Specification



Note: Figure displays estimates of (a) the specific majors' earnings premium, (b) earnings growth from the previous period for general majors, (c) earnings growth from the previous period for specific majors and (d) the earnings growth of specific majors-general majors. Plotted are the coefficients from three different regression models that vary in the functional form of years since graduation (quadratic, 2 year bins, and 1 year bins). For these regressions, the estimates come from a regression of (log) annual earnings on the function of years since graduation, dummies for college major specificity groups and interactions of the two. Estimates for "earn change" come from a regression of individual-level changes in earnings on two-year bins, college major dummies and interactions of the two (similar to Equation A.1 but with college major dummies instead of job changing dummies). Individual-level changes in earnings is $\Delta(earn_{imt}) = Y_{i,m,t} - Y_{i,m,t-1}$. For the 1-year and 2-year bin model earnings growth is calculated using changes in mean earnings between major x years since graduation cells: equal to $(\beta_g - \beta_{g-1})$ from period $g-1$ to g for general majors and $(\beta_{m,g} - \beta_{m,g-1}) + (\beta_g - \beta_{g-1})$ for specific majors. For the "earn change" model earnings growth is calculated as $\xi_0 + \xi_g$ for general majors and $\xi_0 + \xi_g + \xi_m + \xi_{g,m}$, and the difference in (c) is $\xi_m + \xi_{m,g}$. Data source is the ACS-LEHD, see Section 1.2.2 for details on the analysis sample. All regression includes 2,637,000 observations for 383,000 individuals, except for "earn change" which includes 2,254,000 observations for 360,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board.

Figure A.2: Differences in Labor Supply between General and Specific Majors



Note: Figure displays estimates of (a) log(labor supply) and (b) the probability of working full-time, full-year. Labor supply is defined as usual weeks worked times usual hours worked per week. Full-time, full-year is defined as 35+hours/week and 40+weeks/year. The level of the outcome for general majors in bin g is calculated as $\beta_0 + \beta_g$ where β_0 is the sample mean of the outcome among general majors in the omitted period and for specific majors is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. All regression include controls as described in Section 1.3. Data source is the 2009-2019 public-use ACS. Sample includes all employed four-year college graduates, that are from the 1999-2012 graduating cohorts, 1-15 years post graduation. Estimates are weighted using ACS survey weights.

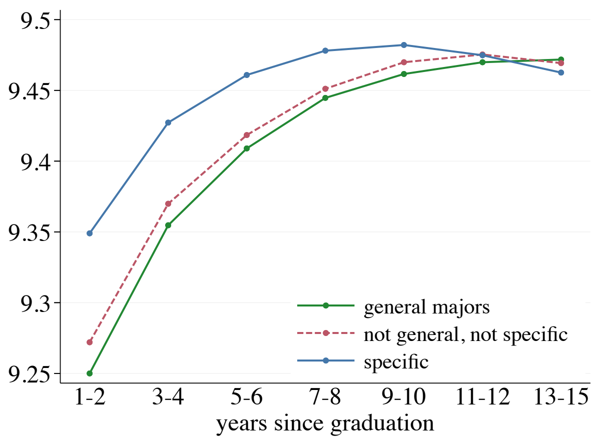
Table A.12: Linear Measure of College Major Specificity: Differences from the Mean

	3-4	5-6	7-8	9-10	11-12	13-15
	Panel A: Log Annual Earnings					
mean + 5	0.0158***	0.0120***	0.0092***	0.0068***	0.0049***	0.0030***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0006)
mean + 10	0.0316***	0.0239***	0.0183***	0.0136***	0.0099***	0.0059***
	(0.0005)	(0.0005)	(0.0006)	(0.0007)	(0.0009)	(0.0012)
mean + 15	0.0474***	0.0359***	0.0275***	0.0204***	0.0148***	0.0089***
	(0.0007)	(0.0008)	(0.0009)	(0.0010)	(0.0013)	(0.0018)
	Panel B: Broad Occupation Change (Weighted)					
mean + 5	-0.0231***	-0.0232***	-0.0192***	-0.0132***	-0.0136***	-0.0142***
	(0.0037)	(0.0031)	(0.0029)	(0.0028)	(0.0029)	(0.0027)
mean + 10	-0.0462***	-0.0464***	-0.0384***	-0.0265***	-0.0272***	-0.0284***
	(0.0074)	(0.0061)	(0.0057)	(0.0057)	(0.0059)	(0.0055)
mean + 15	-0.0693***	-0.0695***	-0.0576***	-0.0397***	-0.0408***	-0.0426***
	(0.0111)	(0.0092)	(0.0086)	(0.0085)	(0.0088)	(0.0082)
	Panel C: Employer Change					
mean + 5	-0.0057***	-0.0056***	-0.0050***	-0.0045***	-0.0042***	-0.0031***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
mean + 10	-0.0114***	-0.0113***	-0.0101***	-0.0089***	-0.0083***	-0.0063***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
mean + 15	-0.0171***	-0.0169***	-0.0151***	-0.0134***	-0.0125***	-0.0094***
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0006)	(0.0007)
	Panel D: Broad Industry Change					
mean + 5	-0.0067***	-0.0058***	-0.0049***	-0.0042***	-0.0037***	-0.0031***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
mean + 10	-0.0133***	-0.0116***	-0.0098***	-0.0083***	-0.0074***	-0.0063***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0004)
mean + 15	-0.0200***	-0.0174***	-0.0147***	-0.0125***	-0.0110***	-0.0094***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0005)

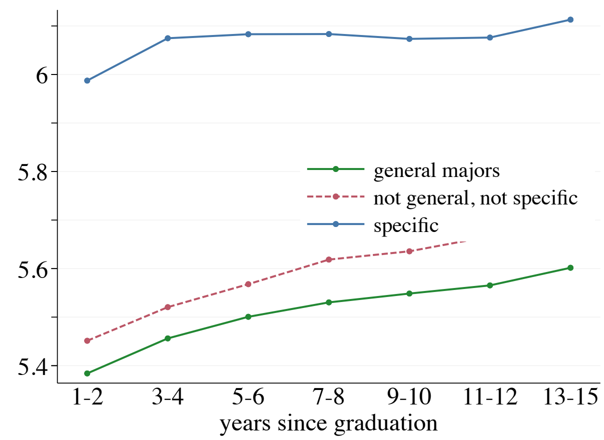
Note: Table displays point estimates for (a) log annual earnings, (b) broad occupation changes, (c) employer changes and (d) indusry changes from a regression of the outcome on a de-meaned linear measure of college major specificity, two-year since graduation bins and interaction of the two. The linear measure of occupational dispersion is the percent of a major's recent graduates in the major's three largest. I subtract the mean across majors of .317. For each year since graduation, table provides estimates of $\beta_0+\beta_g+(z/100)\beta_{occpct}+(z/100)\beta_{occpct,g}$, where β_{occpct} is the coefficient on the de-meaned linear occupational dispersion measure, $\beta_{occpct,g}$ is the coefficient on the interaction between the measure and years since graduation bin, and z is the value given by $mean{+}z$. For example, the point estimates in rows corresponding the $mean + 10$ come correspond to the differential between a major with the average value of the measure ($\beta_{occpct} = 0$) and a major with a value that is 10 percentage points above the mean ($\beta_{occpct} = .1$). Data source for Panels A, C and D is the ACS-LEHD and for Panel B is the ACS-NSCG, see Section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Panel A includes 2,637,000 observations for 383,000 individuals, Panel C and D includes 2,254,000 observations for 360,000 individuals and Panel D includes 61,500 observations for 35,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded. *p< 0.1, **p<0.05, ***p<0.01.

Figure A.3: Employer Attributes & Cumulative Employers

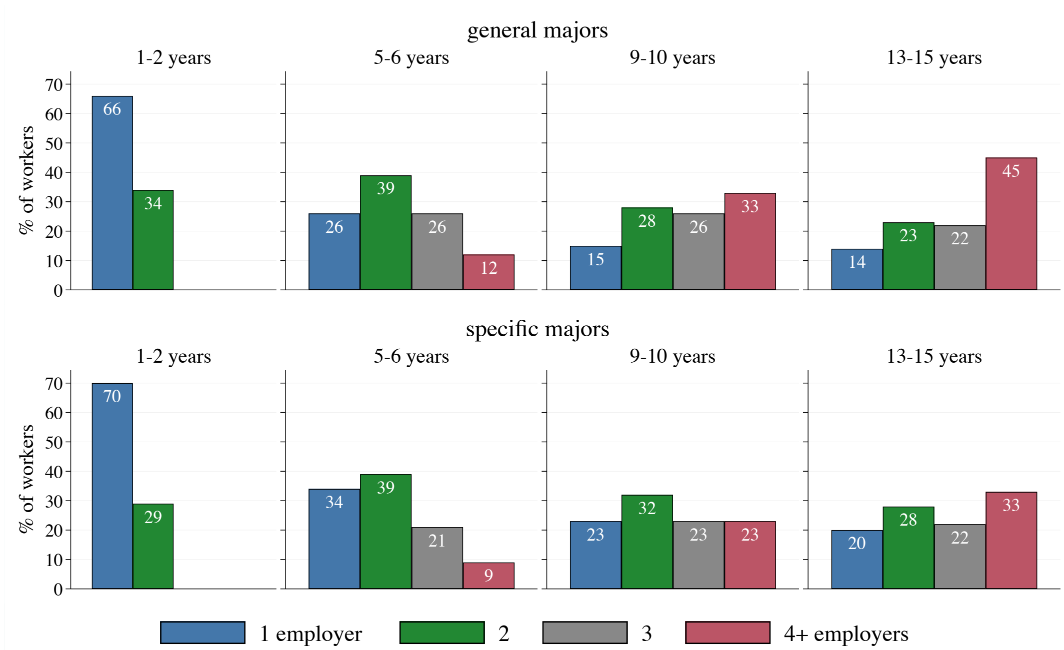
(a) Log Employer Pay



(b) Log Employer Size



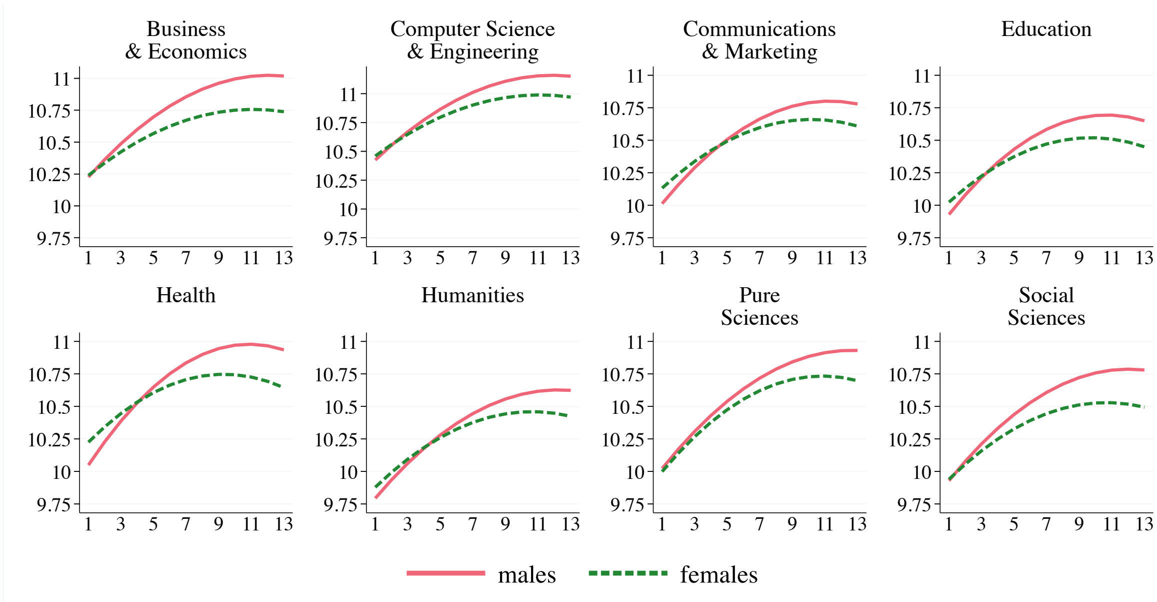
(c) Cumulative Employers



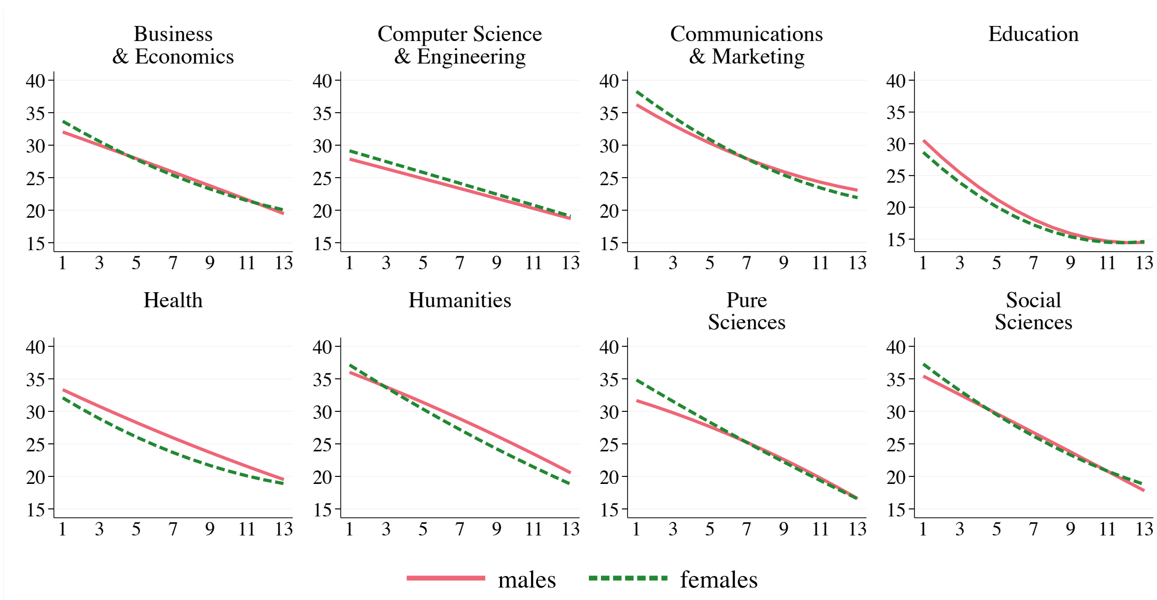
Note: Figure displays (a) mean log employer pay, (b) log employer size and (c) the cumulative number of employers seperately by college major group and years since graduation bin. Coefficients are from a regression of the outcome on two-year bins for years post graduation, dummies for college major specificity groups and interactions of the two (see Equation (1.3.1)). College major specificity groups include general, specific and not general or specific majors (see Figure 1.1). The outcome for general majors in bin g is $\beta_0 + \beta_g$ where β_0 is the sample mean among general majors in the omitted period and for specific majors is $\beta_0 + \beta_g + \phi_m + \beta_{m,g}$. All regression include controls as described in Section 1.3. Outcomes in Panel (a) and (b) are employer characteristics from the restricted-used Quarterly Workforce Indicators (QWI) file and are quarterly measures at the SEINUNIT (establishment) level which I convert to annual measures at the SEIN (state-level employer) level. Employer size is the average (across quarters) number of full-quarter employees and employer pay is the average (across quarters) of earnings per full-quarter employee. The variables are measured in the first year of the employer-employee relationship and are fixed until separation. The outcome in Panel (c) is the cumulative number of employers. Data source is the ACS-LEHD and includes 2,254,000 observations for 360,000 individuals. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

Figure A.4: Gender Differences in Specific-General Major Earnings Gap

(a) Log Annual Earnings



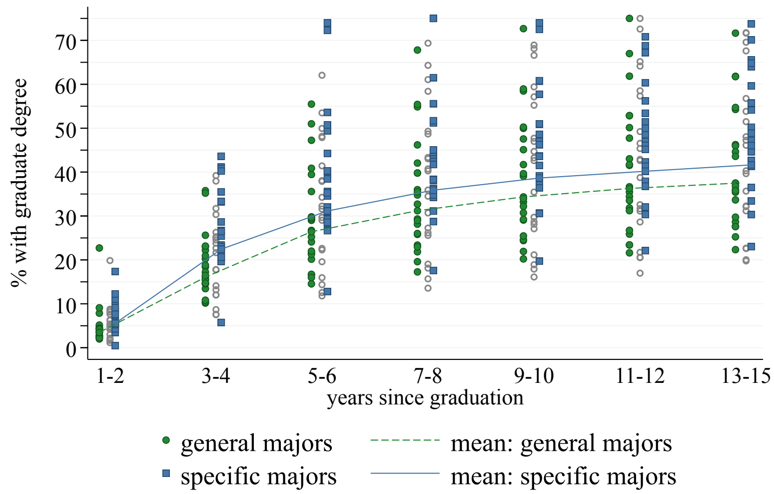
(b) Employer Changes



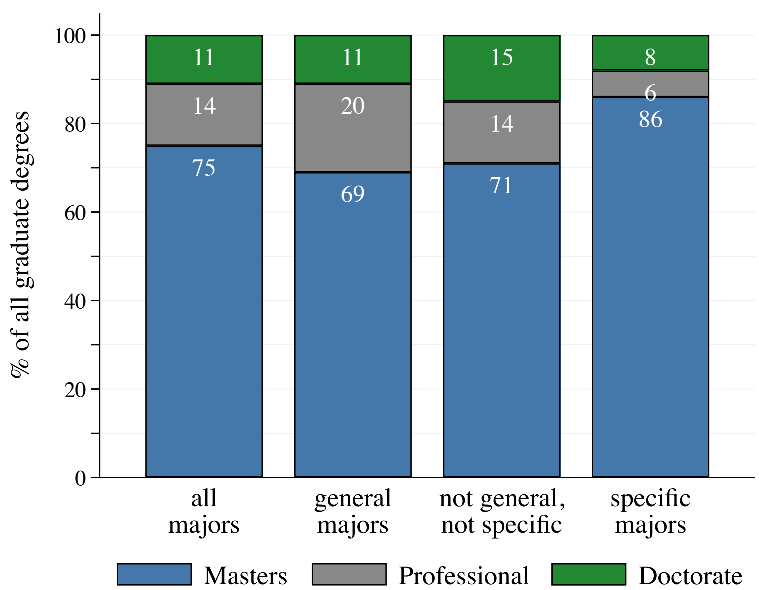
Note: Figure displays estimates of (a) log annual earnings and (b) employer changes for eight college major subject fields and separately by gender. Coefficients are from gender-specific regressions of the outcome on a quadratic function of years since graduation, dummies for college major subject field and interactions of the two (see Equation (1.3.1)). Majors are grouped into nine mutually exclusive college major subject fields and Education is the omitted major (see Appendix Table A.16). Estimates for "All Other Majors" are omitted from the graph. In Panels (a)-(c) the outcome for female (male) Education majors in bin g is $\beta_0 + g\beta_g + g^2\beta_{g2}$ where β_0 is the sample mean among female (male) general majors in the omitted period and for female (male) college major subject field m is $\beta_0 + \phi_m + g(\beta_g + \beta_{g,m}) + g^2(\beta_{g2} + \beta_{g2,m})$. Data source for all panels is the ACS-LEHD. In Panel (a) regression includes 1,557,000 female and 1,080,000 male observations. In Panel (b) regression 1,329,000 female and 926,000 male observations. Section 1.2.2 for details on the analysis sample. All regression include controls as described in Section 1.3. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

Figure A.5: Graduate Degree Attainment

(a) Attainment Levels Over the Early Career



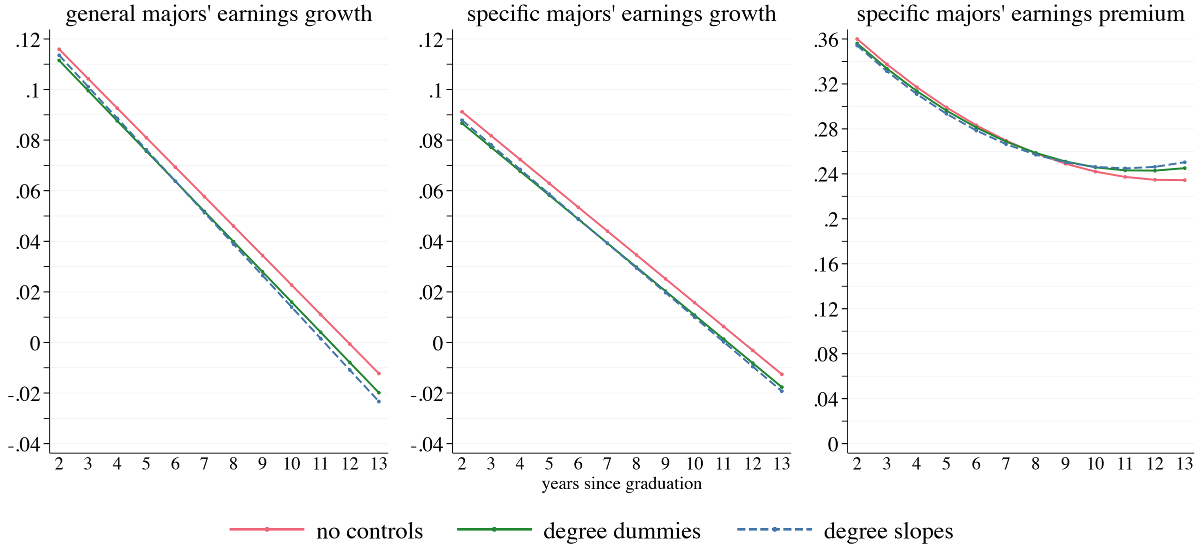
(b) Graduate Degree Type by 13-15 Years Post Graduation



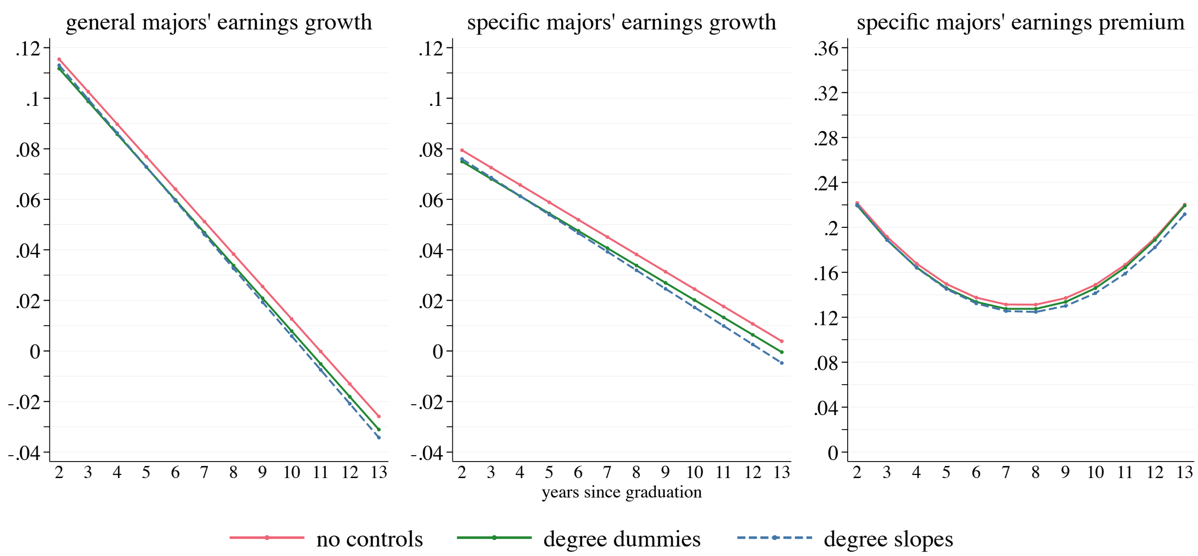
Note: Figure plots graduate degree attainment levels over the early career and separately for college majors. Data source is the public-use 2009-2019 ACS. Panel (a) plots the percent of college graduates in each college major that have a graduate degree. There is one data point for each of the 61 majors. College majors are clustered into three groups including general, specific and not general or specific majors using the distribution of each major's employment across occupations. Panel (b) plots the percent of individuals that are 13-15 years post graduation with graduate degrees that have each graduate degree type (Master's, Professional, Doctorate).

Figure A.6: Sensitivity of Earnings Growth and Gaps to Graduate Education Controls

(a) Unweighted Regression



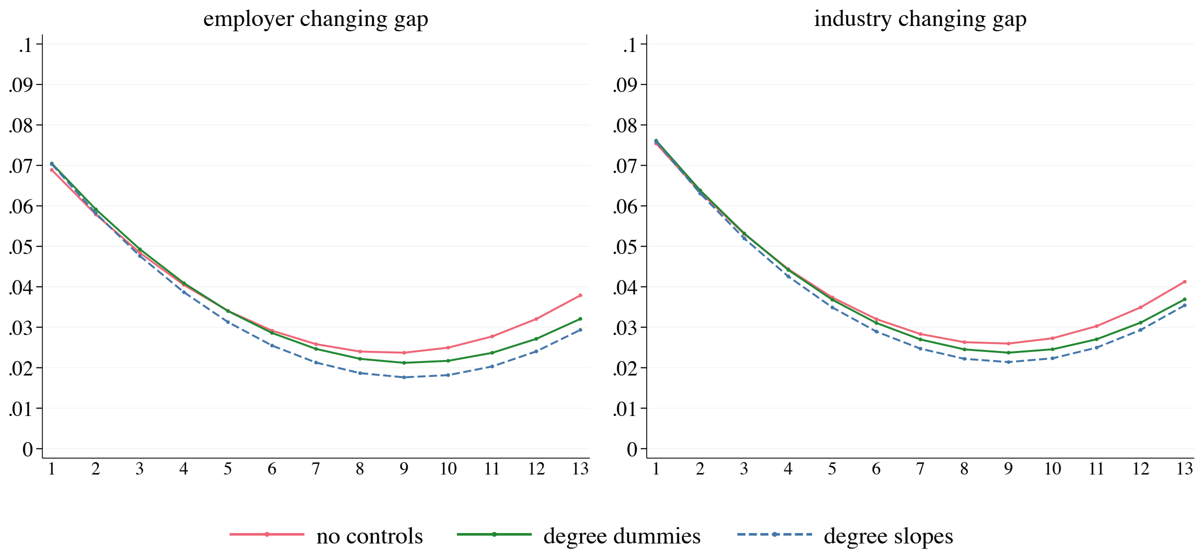
(b) Weighted Regression



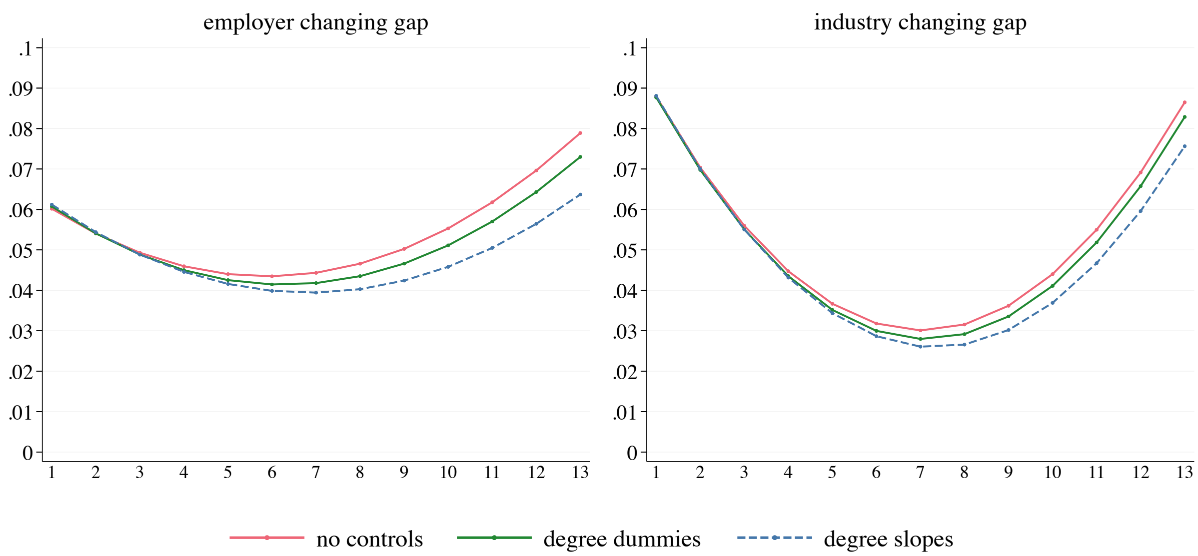
Note: Figure displays estimates of general majors' earnings growth, specific majors' earnings growth and the specific majors' earnings premium and for three different regression specifications. For each outcome, estimates for "no controls" are from a regression of the outcome on a quadratic in years since graduation, college major specificity group, interactions of the two; estimates for "degree dummies" add in a total of 6 graduate degree-type indicators for enrollment and attainment (Master's, Professional, Doctorate): $enroll_{itd}$ and $attain_{itd}$; estimates for "degree slopes" add in the enrollment and attainment slopes: $enroll\ slope_{itd}$ and $attain\ slope_{itd}$. See Equation 1.6.1. Earnings growth for general majors in year g is equal to $[g-(g-1)]\beta_g + [g^2-(g-1)^2]\beta_{g2}$ and for specific majors $[g-(g-1)](\beta_g+\beta_{g,m}) + [g^2-(g-1)^2](\beta_{g2}+\beta_{g2,m})$. The specific majors' earnings premium in year g is equal to $[\phi_m + g(\beta_{g,m}) + g^2(\beta_{g2,m})]$. All regression include controls as described in Section 1.3. Data source is the NSCG-LEHD. See section 1.2.2 for details on the analysis sample. Panel (b) estimates are weighted using NSCG sample weights. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. See Table 1.2 for the regression results.

Figure A.7: Sensitivity of Differences in Job Changing between General and Specific Majors to Graduate Education Controls

(a) Unweighted Regression



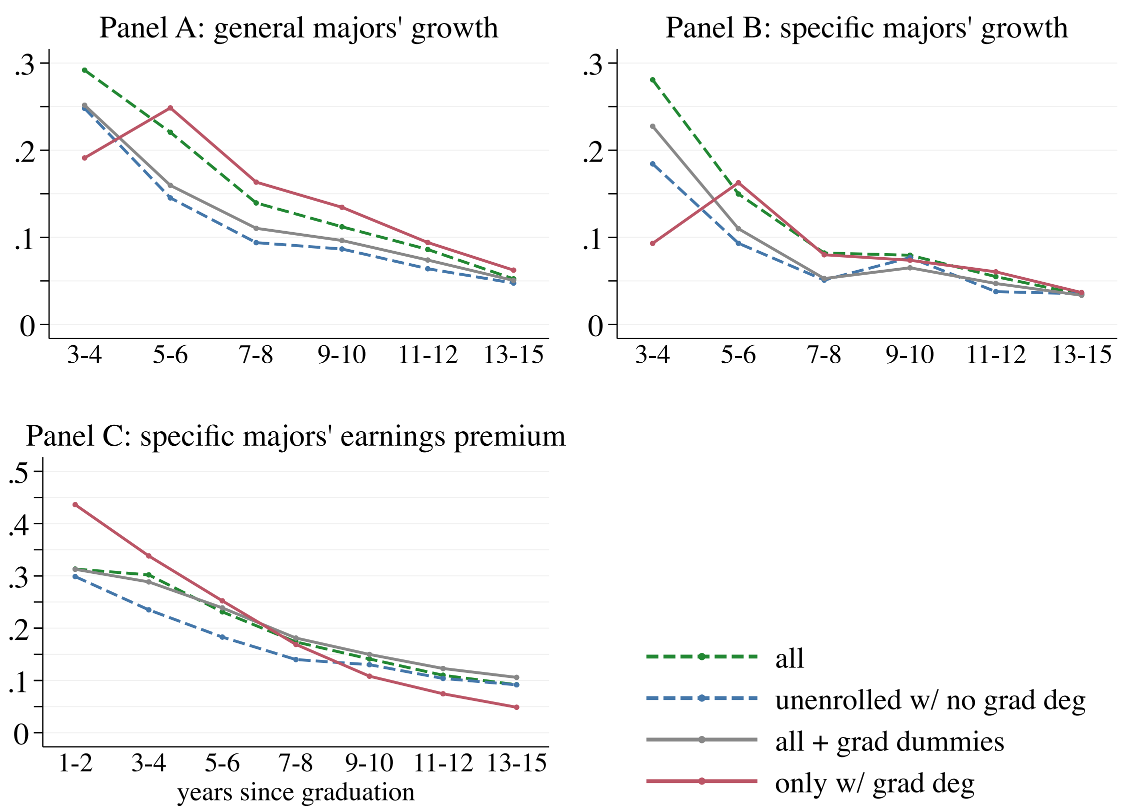
(b) Weighted Regression



Note: Figure displays estimates of the different between general and specific majors in employer changing and industry changing and for three different regression specifications. For each outcome, estimates for "no controls" are from a regression of the outcome on a quadratic in years since graduation, college major specificity group, interactions of the two; estimates for "degree dummies" add in a total of 6 graduate degree-type indicators for enrollment and attainment (Master's, Professional, Doctorate): $enroll_{itd}$ and $attain_{itd}$; estimates for "degree slopes" add in the enrollment and attainment slopes: $enroll\ slope_{itd}$ and $attain\ slope_{itd}$. See Equation 1.6.1. The employer (or industry) changing gap in year g is equal to $-[\phi_m + g(\beta_{g,m}) + g^2(\beta_{g2,m})]$. All regression include controls as described in Section 1.3. Data source is the NSCG-LEHD. See section 1.2.2 for details on the analysis sample. Panel (b) estimates are weighted using NSCG sample weights. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. See Table 1.2 for the regression results.

Figure A.8: Earnings Growth and Gaps in the ACS with and without Graduate Education

(a) Unweighted Regression



Note: Figure displays estimates of (a) general majors' earnings growth, (b) specific majors' earnings growth and (c) the specific majors' earnings premium for four different samples. Sample for "all" include all non-zero earners, sample for "unenrolled w/ no grad deg" drops non-zero earnings who are enrolled and have a graduate degree. Sample for "all + grad dummies" includes all non-zero earners and dummy variables for enrollment and attainment of each graduate degree type (Master's, Professional, Doctorate). Sample for "only w/ grad deg" includes all non-zero earners that have obtained a graduate degree. Earnings growth is calculated using changes in mean earnings between major x years since graduation cells: equal to $(\beta_g - \beta_{g-1})$ from period $g-1$ to g for general majors and $(\beta_{m,g} - \beta_{m,g-1}) + (\beta_g - \beta_{g-1})$ for specific majors. The specific major earnings premium is $\beta_{m,g} + \phi_m$. All regression include controls as described in Section 1.3. Data source is the 2009-2019 public-use ACS. Sample includes all employed four-year college graduates, that are from the 1999-2012 graduating cohorts, 1-15 years post graduation. Estimates are weighted using ACS survey weights.

Table A.13: Regression Coefficients: Accounting for Earnings Growth Differences between Majors, an Oaxaca-Blinder Decomposition

years post graduation	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Unadjusted earnings growth:							
general majors	0.1930***	0.1159***	0.0892***	0.0632***	0.0458***	0.0418***	0.0457***
	(0.0014)	(0.0010)	(0.0011)	(0.0011)	(0.0012)	(0.0015)	(0.0018)
specific majors	0.2274***	0.1104***	0.0686***	0.0499***	0.0346***	0.0330***	0.0392***
	(0.0017)	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0015)	(0.0018)
gap: general - specific	-0.0344***	0.0055***	0.0206***	0.0133***	0.0112***	0.0088***	0.0065**
	(0.0022)	(0.0015)	(0.0016)	(0.0016)	(0.0018)	(0.0022)	(0.0026)
Explained:							
total	0.0055***	0.0104***	0.0120***	0.0072***	0.0051***	0.0042***	0.0002
	(0.0008)	(0.0006)	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0009)
demographics	0.0017***	0.0021***	0.0020***	0.0020***	0.0020***	0.0017***	0.0014***
	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
economic	-0.0091***	-0.0008	0.0045***	0.0016**	0.0013**	0.0012	-0.0024***
	(0.0007)	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0008)	(0.0009)
job changing	0.0129***	0.0091***	0.0055***	0.0036***	0.0018***	0.0013***	0.0012***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Unexplained:							
total	-0.0399***	-0.0049***	0.0086***	0.0061***	0.0061***	0.0046**	0.0063**
	(0.0021)	(0.0015)	(0.0015)	(0.0016)	(0.0017)	(0.0020)	(0.0025)
demographics	-0.0076	-0.0564***	0.0024	-0.0074	-0.0221***	-0.0035	-0.0076**
	(0.0118)	(0.0095)	(0.0052)	(0.0057)	(0.0073)	(0.0025)	(0.0030)
economic	-0.0355	-0.0360	0.0214	0.0498	0.0009	-0.0330	-0.0153
	(0.0298)	(0.0226)	(0.0303)	(0.0328)	(0.0327)	(0.0442)	(0.0445)
job changing	-0.0139***	-0.0080***	-0.0016	-0.0014	-0.0011	-0.0013	00
	(0.0017)	(0.0011)	(0.0011)	(0.0012)	(0.0013)	(0.0015)	(0.0018)
Observations	246,000	416,000	342,000	272,000	202,000	136,000	89,000

Note: Table provides estimates from separate Oaxaca-blinder style decompositions of the earnings-growth difference between general and specific majors in each year since graduation. Each cell provides results from a separate regression. Sample is the ACS-LEHD. Earnings growth is defined as $\Delta(log(earn_{imt})) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$ and is the unadjusted (raw) individual-level earnings growth, measured by year-on-year changes in log earnings. The unadjusted (raw) average among general (*gen*) and specific majors (*spec*) are equal to $\overline{\Delta(log(earn_i))}_{gen}$ and $\overline{\Delta(log(earn_i))}_{spec}$, respectively. The unadjusted earnings-growth difference is just the difference in means: $\overline{\Delta(log(earn_i))}_{gen} - \overline{\Delta(log(earn_i))}_{spec}$. For 3 different groups of covariates, the table displays estimates of the share of the total explained earnings-growth difference accounted for by only the group of covariates: for a given X_j this is $\left(\frac{\beta_{j,P}(\overline{X}_{j,gen} - \overline{X}_{j,spec})}{\sum_k \beta_{k,P}(\overline{X}_{k,gen} - \overline{X}_{k,spec})}\right)$. The covariate groups include (1) individual-level attributes (indicators for female, black, hispanic, and five-year bins for college graduation cohort), (2) a set of economic controls (de-meaned unemployment rate at graduation, state and year fixed effects and a part-year employment indicator) and (3) job change indicators (for employer and industry changes). For a given years since graduation bin, the analysis sample observations from the ACS-LEHD and for general and specific majors. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

A.1.1 Earnings Changes with Job Changes

In this section I estimate whether, on average, year-on-year earnings changes are larger or smaller among workers who changed jobs relative to those who stayed in their job. Second, I estimate whether, on average, the earnings gains of job changers (relative to job stayers) differs between general and specific majors. Together these results will indicate whether the higher earnings growth of general majors is only accounted for by more frequent job changes, or whether it is also accounted for by higher earnings gains at job changes relative to specific majors.

Methods I first estimate differences in average year-on-year earnings gains between job changers and stayers with the following regression specification:

$$\Delta(earn_{imt}) = \xi_0 + \sum_g \xi_g g_{it} + \xi_{djob}(\text{job}\Delta_{it}) + \sum_g \xi_{g,djob}(g_{it} \cdot \text{job}\Delta_{it}) + \delta X_{it} + \psi Z_i + \theta_t + \gamma_{it} + \epsilon_{it} \tag{A.1}$$

where $\Delta(earn_{imt}) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$ is the individual-level change in log earnings between year $t-1$ and year t. The variable g_{it} indexes years since (undergraduate) graduation. The variables γ_{it} and θ_t are state of employment and year fixed effects, Z_i are individual characteristics, and X_{it} is a vector of time-varying individual covariates (see Section 1.3 for a full description of the covariates). The variable $job\Delta_{it}$ is an indicator for whether or not the individual changed jobs from the previous year. A job change can occur in the worker changed only employer or both employer and industry.

Equation (A.1) indexes years since graduation using two-year bins. The change in earnings change in period g from the previous period $g-1$ is equivalent to $\xi_0 + \xi_g$ for job stayers and is equal to $\xi_0 + \xi_g + \xi_{djob} + \xi_{g,djob}$ for job changers. It measures the (weighted) average of year-on-year earnings gains over several periods. Specifically, if bin g spans years since graduation c and d, then $\xi_0 + \xi_g$ is equal to the weighted average of $(earn_d - earn_c)$ and $(earn_c - earn_b)$. Consequently, earnings-growth estimates from Equation (A.1) will differ from those yielded from the earnings level model Equation (1.3.1) with two-year bins, but will be similar to estimates from the earnings level model with one-year bins.[1]

[1]The earnings-growth estimates will also differ for two additional reasons. First, demographic and economic controls serve different functions in a model when the outcome is earnings *changes* rather than earnings *levels*. Second, as the outcome is individual-level changes in earnings, and the panel is not completely balanced panel (i.e. workers move in and out of the analysis sample), mean differences are not equal to differences in means. Earnings-growth estimates from two separate one-year bin models, one with earnings *changes* and the other with earnings *levels* as the outcome, will only be equivalent if the data is a completely balanced panel (no attrition) and there are no controls.

The primary purpose of Equation (A.1) is to test for differences in mean earnings gains between job stayers and changers. Year-on-year earnings gains are higher among job changers than stayers in the base period if $\xi_{djob} > 0$ and in period g if $\xi_{djob} + \xi_{djob,g} > 0$. Equation (A.1) assumes that the average year-on-year earnings gains of job changers (and of job stayers) are equivalent for general and specific majors. I relax this assumption by estimating the following model:

$$\begin{aligned}\Delta(earn_{imt}) = \beta_0 + \sum_g \beta_g g_{it} + \beta_{djob}(\text{job}\Delta_{it}) + \sum_m \beta_m(\text{major group}_{im}) + \sum_g \beta_{g,djob}(g_{it} \cdot \text{job}\Delta_{it}) \\ + \sum_g \sum_m \beta_{g,m}(g_{it} \cdot \text{major group}_{im}) + \sum_g \sum_m \beta_{djob,m}(\text{major group}_{im} \cdot \text{job}\Delta_{it}) \\ + \sum_g \sum_m \beta_{djob,m,g}(g_{it} \cdot \text{major group}_{im} \cdot \text{djob}\Delta_{it}) + \delta X_{it} + \psi Z_i + \theta_t + \gamma_{it} + \epsilon_{it}\end{aligned} \tag{A.2}$$

where $job\Delta_{it}$ and g_{it} are as before, but now there are interactions with college major specificity group $major\ group_{im}$. Equation (A.2) allows for two forms of earnings gains heterogeneity. First, it allows earnings gains to differ between job changers and stayers *conditional on major*. Among workers with major m, the earnings gains for job changers exceed those of job stayers if $\beta_{djob} + \beta_{djob,m} + \beta_{djob,g} + \beta_{djob,m,g} > 0$ (and for the omitted major if $\beta_{djob} + \beta_{djob,g} > 0$). Second, it allows the earnings gains of job changers (and of job stayers) to differ between general and specific majors. The earnings gains of job changers are higher for major m than for the omitted major if $\beta_m + \beta_{djob,m} + \beta_{m,g} + \beta_{djob,m,g} > 0$. Similarly, the earnings gains of job stayers are higher for major m than for the omitted major if $\beta_m + \beta_{m,g} > 0$.

Results Figure A.9 presents estimates of Equation (A.1) and (A.2) with log annual earnings changes as the outcome using the ACS-LEHD sample and years since graduation indexed by two-year bins. See Appendix Tables A.14 and A.15 for full regression results.

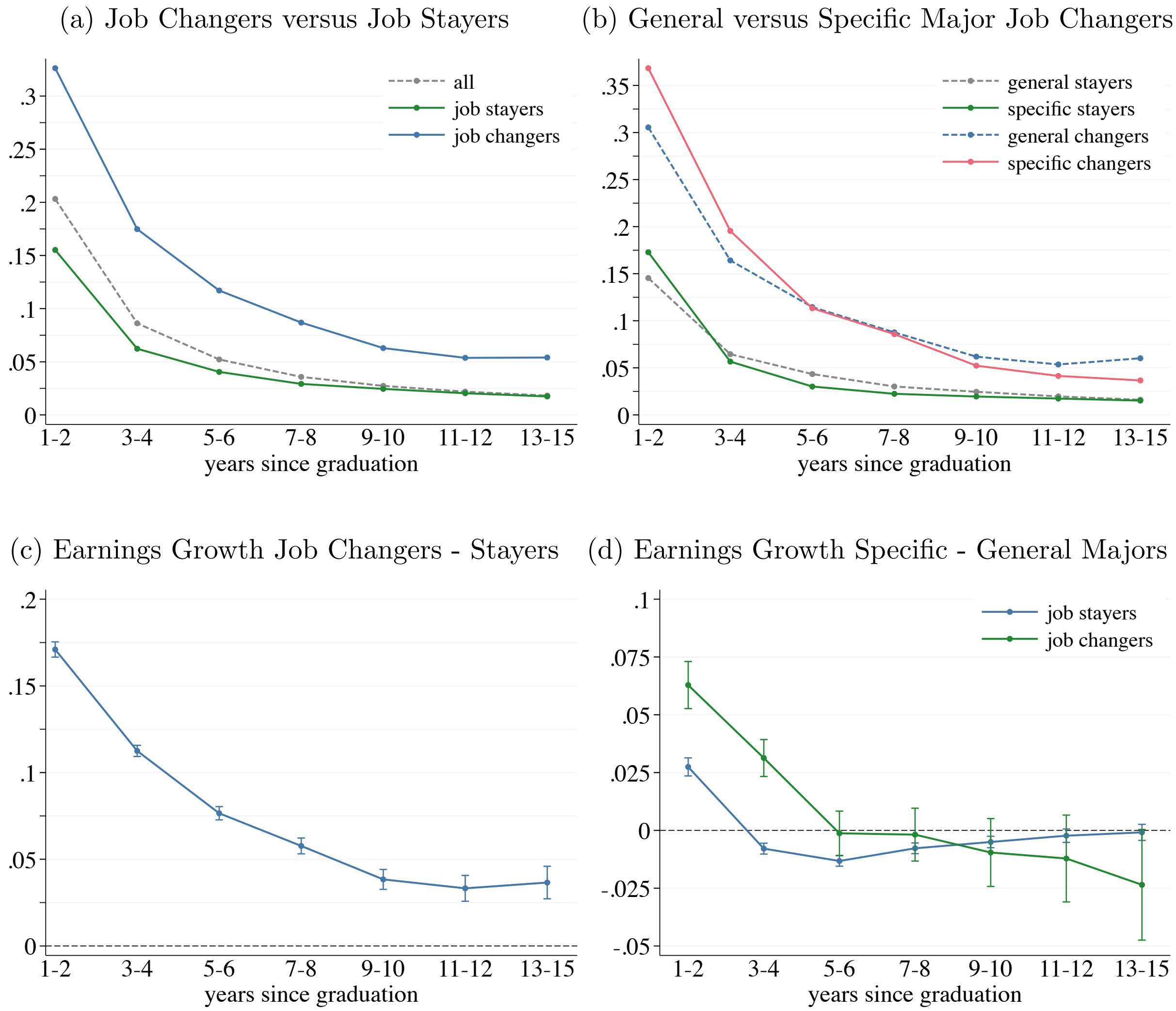
The average year-on-year earnings gains of workers who changed jobs far exceed the gains among workers who did not (see Figure A.9a and A.9c). The average earnings growth (i.e. change in log earnings) among workers who are 3-4 years post graduation was 8%, but among job changers the gains are 17% which is more than double the average of 6% among job stayers. Average earnings gains of all workers steadily decrease over the early career to around 2% at the end of the analysis window, but the differential between job changers and stayers stays statistically different from zero and ranges from 3-7.5 (p<.01) percentage points.

Figure A.9d shows that there are some between-major differences in the earnings growth among job stayers and among job changers. Generally, the between-major differences are larger earlier in the career, and are smaller among job stayers than among job changers. Specifically, in the first two years post graduation, the average earnings growth of specific majors that didn't change jobs exceeds that of general majors by 3 percentage points (p<.01), but in subsequent periods, the earnings growth among general majors tend to be slightly higher (<1 percentage point). Early only, specific majors that changed jobs experience larger earnings growth than general majors that changed jobs – a difference of 6 and 3 percentage points (p<.01) in the periods 1-2 and 3-4 years post graduation – but in the following years, the between-major differences in earnings growth among job changers are small and not statistically different from zero.

Figure A.9b, however, illustrates that the gap in earnings growth between job stayers and changers *conditional on major* are far larger than those between general and specific majors *conditional on job changing status*. Among general majors, the gap in earnings growth between job changers and stayers is initially 16 percentage points and among specific majors the figure is 19 percentage point. In subsequent periods, the difference ranges from around 4 to 10 percentage points among general majors and from 2-14 percentage points among specific majors. These differences far exceed the between-major differences among job stayers (<3%) and among job changers (<6%) discussed in the previous paragraph.

This pattern of results suggests that the higher earnings growth of general majors is not primarily accounted for by higher earnings gains of job changers relative to specific majors. Rather, they are primarily the result of more frequent job changes and large differences in the mean earnings growth of job changers relative to stayers irrespective of major type.

Figure A.9: Earnings Changes and Job Changes



Note: Figure displays estimates of individual-level changes in earnings (a) separately for job changers and job stayers and (b) separately for general majors that are job stayers, general majors that are job changers, specific majors that are job stayers, and specific majors that are job changers. Panels (c) plots the difference in earnings changes between job changers and job stayers from panel (a). Panel (d) plots the difference in earnings changes between general and specific majors for job stayers, and then separately for job changers. Estimates in Panel (a) and (c) are from Equation A.1. Earnings changes in (a) for job stayers=$\xi_0 + \xi_g$ and for job changers=$\xi_0 + \xi_g + \xi_{djob} + \xi_{g,djob}$, and the difference in (c) is $\xi_{djob} + \xi_{djob,g}$. Estimates in Panel (b) and (d) are from Equation A.2. Earnings changes in (b) for general stayers=$\beta_0 + \beta_g$, for general changers=$\beta_0 + \beta_g + \beta_{djob} + \beta_{g,djob}$, for specific stayers=$\beta_0 + \beta_g + \beta_m + \beta_{m,g}$, and for specific changers=$\beta_0 + \beta_g + \beta_m + \beta_{m,g} + \beta_{djob} + \beta_{g,djob} + \beta_{djob,m} + \beta g, djob, m$. The difference in (d) for job changers=$\beta_m + \beta_{djob,m} + \beta_{m,g} + \beta_{djob,m,g}$ and for job stayers=$\beta_m + \beta_{m,g}$. The outcome is $\Delta(earn_{imt}) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$, the individual-level change in (log) earnings between year $t-1$ and year t. Job change is defined as a change in employer or change in employer and industry from the previous observation. All regression include controls as described in Section 1.3. Data source is the ACS-LEHD, 2,254,000 observations for 360,000 individuals. See section 1.2.2 for details on the analysis sample. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. Regression results are in Appendix Table A.14 and A.15.

Table A.14: Regression Coefficients: Year-on-Year Individual Earnings Gains

Year-on-Year Individual Earnings Gains: By College Major							
years post grad:	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general	0.193	-.1062***	-.1378***	-.1549***	-.1648***	-.1707***	-.1738***
		(.0017)	(.0019)	(.0021)	(.0023)	(.0028)	(.0037)
not gen/not spec	-.0006	-.0040	-.0004	-.0042	-.0004	-.0014	-.0025
	(.0022)	(.0028)	(.0028)	(.0029)	(.0030)	(.0033)	(.0036)
specific	.0293***	-.0322***	-.0431***	-.0383***	-.0363***	-.0341***	-.0346***
	(.0021)	(.0027)	(.0027)	(.0027)	(.0028)	(.0030)	(.0033)
Panel B: earnings growth from previous period							
general	0.1930	0.0868	0.0552	0.0381	0.0282	0.0223	0.0192
not gen/not spec	0.1924	0.0822	0.0542	0.0334	0.0272	0.0204	0.0161
specific	0.2223	0.0839	0.0414	0.0291	0.0212	0.0175	0.0138
Panel C: earnings growth gap (major m - general)							
not gen/not spec	-.0006	-.0046***	-.0010	-.0047***	-.0010	-.0019	-.0031
	(.0022)	(.0016)	(.0017)	(.0018)	(.0020)	(.0024)	(.0029)
specific	.0293***	-.0029**	-.0138***	-.0090***	-.0070***	-.0048**	-.0054**
	(.0021)	(.0015)	(.0015)	(.0015)	(.0017)	(.0020)	(.0025)
Year-on-Year Individual Earnings Gains: By Job Changers vs Stayers							
Panel A: regression coefficients:							
job stayers	.1552	-.0929***	-.1147***	-.1260***	-.1307***	-.1348***	-.1378***
		(.0011)	(.0013)	(.0015)	(.0018)	(.0023)	(.0032)
job changers	.1710***	-.0585***	-.0944***	-.1133***	-.1326***	-.1377***	-.1344***
	(.0022)	(.0029)	(.0030)	(.0032)	(.0037)	(.0044)	(.0053)
Panel B: earnings growth from previous period							
job stayers	0.1552	0.06229	0.0405	0.0292	0.0245	0.0204	0.0174
job changers	0.3262	0.1748	0.11705	0.0869	0.0629	0.0537	0.054
Panel C: earnings growth gap (job changers-stayers)							
job changers	.1710***	.1125***	.0765***	.0577***	.0384***	.0332***	.0366***
	(.0022)	(.0016)	(.0019)	(.0023)	(.0029)	(.0038)	(.0048)

Note: Top panel displays estimates from a variation of Equation A.1 which is a regression of individual-level changes in log earnings, $\Delta(earn_{imt}) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$ on dummies for college major specificity group, years since graduation bins and interactions of the two. Bottom panel displays estimates from Equation A.1 which is a regression of $\Delta(earn_{imt})$ on dummies for job changers (employer and/or industry change), years since graduation bins and interactions of the two. In bin g, Panel A provides the untransformed estimates and standard errors. In Panel B earnings growth is calculated as: (top) $\xi_0 + \xi_g$ for general majors and $\xi_0 + \xi_g + \xi_m + \xi_{g,m}$ for specific majors, (bottom) $\xi_0 + \xi_g$ for job stayers and $\xi_0 + \xi_g + \xi_{djob} + \xi_{g,djob}$ for job changers. In Panel C the earnings growth gap is calculated as: (top) $\xi_m + \xi_{m,g}$ and (bottom) $\xi_{djob} + \xi_{djob,g}$. All regression include controls as described in Section 1.3. Data source is the ACS-LEHD, 2,254,000 observations for 360,000 individuals. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board.

Table A.15: Regression Coefficients: Year-on-Year Individual Earnings Gains

years post grad:	1-2	3-4	5-6	7-8	9-10	11-12	13-15
Panel A: regression coefficients:							
general, stayers	.1454	-.0808***	-.1020***	-.1152***	-.1208***	-.1257***	-.1293***
		(.0016)	(.0017)	(.0019)	(.0022)	(.0026)	(.0034)
general, changers	.1601***	-.0605***	-.0889***	-.1026***	-.1228***	-.1262***	-.1160***
	(.0032)	(.0042)	(.0044)	(.0048)	(.0055)	(.0067)	(.0082)
not gen/spec, stayers	-.0032	-.0031	.0012	.0007	.0008	.0000	.0009
	(.0021)	(.0025)	(.0025)	(.0025)	(.0025)	(.0027)	(.0029)
not gen/spec, changers	.0110**	-.0032	-.0055	-.0184**	-.0025	-.0027	-.0145
	(.0054)	(.0071)	(.0075)	(.0083)	(.0096)	(.0118)	(.0144)
specific, stayers	.0275***	-.0353***	-.0407***	-.0352***	-.0325***	-.0297***	-.0283***
	(.0020)	(.0023)	(.0023)	(.0023)	(.0024)	(.0025)	(.0027)
specific, changers	.0354***	.0038	-.0234***	-.0295***	-.0400***	-.0453***	-.0581***
	(.0056)	(.0072)	(.0076)	(.0082)	(.0094)	(.0112)	(.0136)
Panel B: earnings growth from previous period							
general, stayers	0.1454	0.0646	0.0434	0.0302	0.0246	0.0197	0.0161
general, changers	0.3055	0.1642	0.1146	0.0877	0.0619	0.0536	0.0602
not gen/spec, stayers	0.1422	0.0583	0.0415	0.0277	0.0222	0.0166	0.0138
not gen/spec, changers	0.3133	0.1656	0.1181	0.0778	0.0680	0.0587	0.0543
specific, stayers	0.1729	0.0567	0.0302	0.0224	0.0196	0.0174	0.0152
specific, changers	0.3684	0.1955	0.1133	0.0859	0.0523	0.0414	0.0367
Panel C: earnings growth gap (job changers-stayers)							
general	0.1601	0.0996	0.0712	0.0575	0.0373	0.0339	0.0441
not gen/spec	0.1711	0.1073	0.0766	0.0501	0.0458	0.0422	0.0406
specific	0.1955	0.1388	0.0831	0.0634	0.0328	0.0240	0.0214
Panel D: earnings growth gap (major m-general majors)							
job stayers	.0275***	-.0079***	-.0132***	-.0078***	-.0050***	-.0023	-.0009
	(.0020)	(.0012)	(.0012)	(.0012)	(.0013)	(.0015)	(.0018)
job changers	.0629***	.0313***	-.0013	-.0018	-.0096	-.0122	-.0235*
	(.0052)	(.0041)	(.0049)	(.0058)	(.0075)	(.0096)	(.0122)

Note: Table displays estimates of Equation A.2 which is a regression of individual-level changes in log earnings, $\Delta(earn_{imt}) = log(Y_{i,m,t}) - log(Y_{i,m,t-1})$, on dummies for job changing, college major specificity group, years since graduation bins and interactions of the three variable types. In bin g, Panel A provides the untransformed estimates and standard errors. Panel B provides earnings growth from previous period for: general, stayers: $\beta_0 + \beta_g$; general, changers: $\beta_0 + \beta_g + \beta_{djob} + \beta_{djob,g}$, specific, stayers: $\beta_0 + \beta_m + \beta_g + \beta_{g,m}$; specific, changers: $\beta_0 + \beta_m + \beta_g + \beta_{g,m} + \beta_{djob} + \beta_{djob,m} + \beta_{djob,g} + \beta_{djob,m,g}$. In Panel C the earnings growth gap between job changers and job stayers for general majors is $\beta_{djob} + \beta_{djob,g}$ and for specific majors is $\beta_{djob} + \beta_{djob,g} + \beta_{djob,m} + \beta_{djob,m,g}$. In Panel D the earnings growth gap between major m and general majors for job changers is $\beta_m + \beta_{djob,m} + \beta_{m,g} + \beta_{djob,m,g}$ and for job stayers is $\beta_m + \beta_{m,g}$. All regression include controls as described in Section 1.3. Data source is the ACS-LEHD, 2,254,000 observations for 360,000 individuals. Standard errors are clustered at the individual level. Results were disclosed by the U.S Census Bureau's Disclosure Review Board.

A.2 Data

A.2.1 College Majors

College Major Codes The most detailed level of major codes I use call into 61 categories. To create the codes I start with the aggregation of 170+ American Community Survey (ACS) codes into 70 majors as done in Hemelt et al (2021). I then perform a few adjustments to use their major scheme in my setting. First, I drop four majors from the Hemelt et al (2021) scheme that don't have corresponding majors in the ACS including Urban Planning, Allied Health, Mental & Social Health Services and Construction Management. Second, there are a few majors that are dissagregated in the Hemelt et al (2021) codes that I aggregate to accommodate the NSCG major codes: I collapse (1) Public Relations, Advertising & Applied Communication and Communication & Media Studies into Communications; and (2) Human Resources Management & Services and Hospitality Administration & Management into Business. For some analysis I also aggregate the detailed majors into 9 mutually exclusive college major subject fields based on the aggregation in Altonji & Zhong 2021. Categories include: Business & Economics, Communications & Marketing, Computer Science & Engineering, Education, Health, Humanities, Pure Sciences (Bio, Life, Physical & Math), Social Sciences, and Other.

Defining College Major Specificity For the vast majority of the empirical analysis, I aggregate the 61 detailed major codes intro three mutually exclusive groups based on a measure of the specificity of a college major's skills to particular occupations: (1) general majors, (2) specific majors, and (3) not general or specific majors. I use the 2009-2019 public-use versions of the American Community Survey (ACS) to measure the percent of recent college graduates in the top three occupations for the major. I first restrict attention to individuals with estimated years of college graduation 1999-2012, who are not enrolled, who are working at least part-time (at least 20 hours a week) and part-year (interval of weeks worked equal to 27-39 weeks or higher), and are in the first five years of the career. I then collapse the data down to the major-occupation level. Occupation is measured using a set of 300+ ACS occupations that I harmonize across ACS survey waves (see more detail below). For each major m, I calculate the percent of all graduates that are employed in each occupation o, $p_{o,m} = \frac{N_{o,m}}{N_m}$ where N_m is the number of workers with major m and $N_{o,m}$ is the number of graduates with major m employed in occupation O. I then sort a major's occupation employment cells from highest to lowest share and sum the shares for top three occupations with the highest three shares : $p_{o=1^{st},m} + p_{o=2^{nd},m} + p_{o=3^{rd},m}$, where $p_{o=1^{st},m}$ is the share of workers with major m that are implied in the largest occupation for that major

$o = 1^{st}$. Majors are then divided into three equally size groups. See full list of college majors and occupational specificity measures in Table A.16.

A.2.2 Creating ACS, NSCG and LEHD Datasets

A.2.2.1 Data Sources

For this study I use three different sources of data including the American Community Survey (ACS), National Survey of College Graduates (NSCG) and the Longitudinal Employer Household Dynamics (LEHD). For the main analysis, all three datasets are accessed via a Federal Statistical Research Data Center (FSRDC) maintained by the U.S. Census Bureau. For some supplementary analyses, I also use the public-use versions of ACS extracted from the Integrated Public Use Micro-data Series (IPUMS) and the NSCG extracted from the Scientists and Engineers Statistical Data System (SESTAT). In what follows I first describe each of the three data sources.

American Community Survey (ACS) The ACS is a survey of U.S. households that was created to replace the Census long form and reaches about 1 in 100 households a year (≈3.5 million households). Each ACS wave is a nationally representative cross-section of the U.S. population (households). In the FSRDC I use the ACS from years 2009-2019.

National Survey of College Graduates (NSCG) The National Survey of College Graduates (NSCG) is sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) and provides data on the characteristics of the nation's college graduates. The survey focuses on workers with at least a bachelor's degree. While the NSCG surveys college graduates in all academic disciplines, there is a particular focus on graduates in the science and engineering workforce.[2]

In each NSCG wave, the sampling frame is drawn from previous respondents to the ACS. Specifically, eligible ACS respondents include those who have at least a bachelor's degree, that are institutionalized and are younger than 76. New respondents in the 2010 NSCG are sampled from 2009 ACS respondents, in the 2013 NSCG from 2011 ACS respondents, in the 2015 NSCG from the 2013 ACS respondents, and so on. In the FSRDC I use the NSCG from years 2010, 2013, 2015, 2017 and 2019.

Starting in 2010 the NSCG implemented a four-panel rotating panel design. In each NSCG wave, the new cohort of individuals (who are selected as described above) receive a

[2]Science and Engineering fields of study include Computer and math sciences, Biological, agricultural, and environmental life sciences, Physical sciences (Physics, chemistry, geosciences), Social Sciences (Psychology, economics, political science) and Engineering.

Table A.16: College Major Subject Fields

College Major Subject Field	College Major
Business & Economics	Accounting Business, general Economics Finance Management Information Systems and Science
Communications & Marketing	Communications Journalism Marketing
Computer Science & Engineering	Aeronautical Engineering Architecture Chemical engineering Civil Engineering Computer and Information Sciences, General Computer Engineering Electrical, Electronics and Communications Engineering Industrial And Manufacturing Engineering Materials Science and Engineering Mechanical engineering Other Engineering
Education	Other Education Special Education and Teaching Teacher Education
Health	Allied Health Health and Medical Administrative Services Nursing Pharmacy Public Health Rehabilitation and Therapeutic Professions
Humanities	Applied Arts English, Liberal Arts, Humanities Foreign Language & Linguistics Other Visual/Performing Arts Philosophy, Religion & Theology
Other	Engineering technology Fitness, Recreation and Leisure Studies Legal Studies Library Science Protective Services
Pure Sciences (Bio, Life, Physical & Math)	Agriculture Atmospheric Sciences and Meteorology Biochemistry, Biophysics and Molecular Biology Biology Biomedical Engineering Chemistry Geology and Earth Science Mathematics Microbiology Natural Resources Nutritional sciences Other Physical Sciences Physics Statistics
Social Sciences	Family and Consumer Sciences Geography Other Social Sciences Political Science, Government, and International Relations Psychology Public Administration Public Policy Social Work Sociology

baseline survey interview and three biennial follow-up interviews. For example, a respondent in the new cohort in 2010 NCSG would be interviewed in the 2013, 2015 and 2017 NSCG. As a result, each NSCG wave contains both new respondents and returning respondents.

The ability to link an individual's responses across NSCG waves is limited in the NSCG public-use files. Specifically, within the public-use files, individuals can only be linked between the 2010, 2013, and 2015 NSCG waves as the unique identifiers are omitted from the public-use files starting in 2017. In addition, linking an individual's responses across the ACS and NSCG is infeasible with the public-use ACS and NSCG. In contrast, in the FSRDC I can link an individual's responses across all NSCG waves and to the ACS.

Longitudinal Employer Household Dynamics (LEHD) The LEHD is a linked employer-employee dataset managed by the U.S. Census Bureau. The data is constructed from the wage records extracted from Unemployment Insurance (UI) administrative files from each LEHD partner state. The wage records cover an individual's UI-covered earning from an employing entity. Included employers are those covered by state unemployment insurance (UI) programs and employers subject to the reporting requirements of the ES-202 system including civilian workers covered by the program of Unemployment Compensation for Federal Employees (UCFE). Overall, the LEHD does not cover the entirety of employment in the U.S but does capture a large portion. The Bureau of Labor Statistics (BLS) reports that in 1991 the UI and UCFE coverage comprised roughly 98% of total wage and salary civilian employment and 94.3% of the wage and salary component of national income. Notable exclusions include some private-sector employment, certain types of nonprofit employers and self-employed agricultural and non-agricultural workers.[3] I hereafter to all employers and employees found in the LEHD data as covered employment.

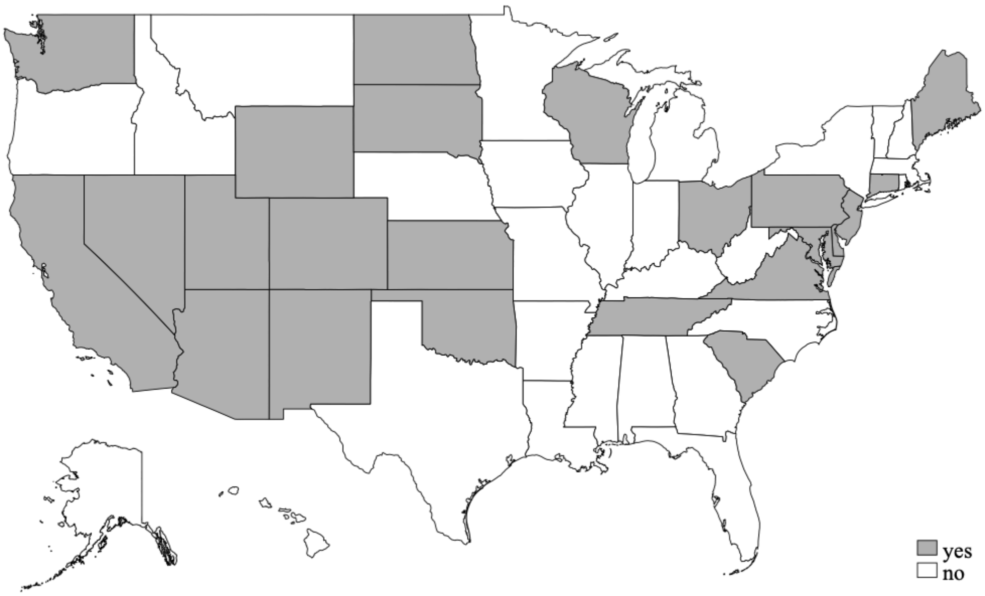
The LEHD infrastructure contains several file types that can be linked together for this analysis including the Employment History Files (EHF) and the Employer Characteristics File (ECF). For each state, the state-specific EHF contains the complete in-state history of employment for all individuals that appear in the UI wage records employed at some firm. An individual will be found in the UI wage records if they have earnings of at least 1 dollar at at least one employer during the quarter. While the full LEHD contains information for all 50 states and D.C. I have access to LEHD records for 23 states (hereafter referred to as covered states). The vast majority of states enter the LEHD system by year 2000 and the historical availability of wage data prior to 2000 differs significantly across states.

A worker's earnings records can be linked across employers and states. In addition,

[3]For example, the BLS also states that in 1991 exclusions amounted to approximately 0.3 million wage and salary agricultural employees, 1.5 million self-employed farmers, 8.9 million self-employed non agricultural workers, 0.7 million domestic workers, and 0.3 million unpaid family workers.

an individuals' records in the LEHD can be linked to other datasets within the FSRDC, including the ACS and the NSCG, via the Census Bureau's unique identifier called the protected identification key or *PIK* and Internal Census crosswalks.

Figure A.10: Coverage of LEHD States



Note: Figure plots the 23 states of data that contribute to the LEHD data used in this article. These states are referred to as covered states.

A.2.2.2 Linking Individuals using Identifiers

There are three different types of individual identifiers key to these linkages: (1) the Census Bureau's unique identifier called the protected identification key or *PIK*; (2) ACS household and individual identifiers, *cmid* and *pnum*; (3) the NSCG unique identifiers of refid. Internal Census crosswalks enable a linkage of *cmid-pnum* to *PIK*. Internal Census datasets also enable a linkage of *refid* to *cmid-pnum*. As all *refid* in the NSCG can be linked to the ACS *cmid-pnum* identifier pairs, *refid* can also be linked to PIK using the *cmid-pnum* to *PIK* crosswalk. Essentially, I am using the ACS as a bridge file to enable the linkage of the NSCG to other datasets that contain PIK. To my knowledge I am the first to use the *cmid-pnum* to PIK crosswalk to link the ACS to the NSCG and the NSCG to the LEHD.

A.2.2.3 Overview of Linked Datasets

Using the identifiers described above, I perform multiple linkages using the three datasets which I summarize below:

- **ACS-NSCG:** I link together an individual's responses across the four waves of the NSCG using *refid*. This yields a short panel of 1-4 observations for NSCG respondents. I then link an individual's NSCG responses to their ACS observation using the *refid* to *cmid-pnum* crosswalk. The final result is a panel of 2-5 observations per individual

- **NSCG-LEHD** I first link together an individual's LEHD observations across years, states and employers using PIK. I link individual-level characteristics from the NSCG to an individual's LEHD observations using the *refid* to *cmid-pnum* to *PIK* mapping. After the data are collapsed to one observation per person per year, the dataset has 1-15 annual observations per individual.

- **ACS-LEHD** This dataset is created in a very similar fashion to the NSCG-LEHD. I then link individual-level characteristics from the ACS to an individual's LEHD observations using PIK (and the *cmid-pnum* to *PIK* mapping). After the data are collapsed to one observation per person per year, the dataset has 1-15 annual observations per individual.

  Note that I take steps to make individuals in the ACS-LEHD and the NSCG-LEHD samples **mutually exclusive**. All individuals in the NSCG, by design, were once surveyed in the ACS. The reverse is not true. For the subject of individuals that were surveyed in both the ACS and NSCG, I only include these individuals and their observations in the NSCG-LEHD sample and not the ACS-LEHD sample. This is done to facilitate disclosure review.

A.2.2.4 Creating ACS-NSCG Master Sample

To create the analysis samples of <u>individuals</u> that will be included in the ACS-NSCG, NSCG-LEHD and ACS-LEHD panels, I proceed in 3 steps: (1) I create a list of *all* individuals from the 2009-2019 ACS with at least a bachelor's degree at the time of the survey; (2) I take steps to deduplicate the list; (3) I impose education restrictions; (4) I impose employment restrictions. I provide details on each step below.

1. **Creating the pool of all individuals** I create a list of all survey respondents from the 2009-2019 ACS with at least a bachelor's degree at the time of the survey. I create a list of all survey respondents from the 2010-2019 NSCG. As described above, link each ACS individual to a *PIK* (using *cmid-pnum* to *PIK* mapping), and each NSCG individual to a *PIK* (using the *refid* to *cmid-pnum* to *PIK* mapping).

2. **Deduplication** I next de-duplicate the list of identifiers to create a unique mapping of *cmid-pnum* to PIK. I drop *cmid-pnum* observations that cannot be linked to a

valid PIK. These observations were flagged and retained in the previous steps when *cmid-pnum* was merged onto the internal *cmid-pnum* to PIK crosswalk. I then drop all *cmid-pnum* observations for which 3 different *cmid-pnum* map to a single PIK. For pairs of *cmid-pnum* that map to a single PIK, I check that the birthdate and place of birth for the two *cmid-pnum* observations match and then randomly keep only one of the *cmid-pnum* to PIK links. Finally in cases where two different *cmid-pnum* link to a single PIK and birthdate and place of birth don't match across the observations, I drop both *cmid-pnum* to PIK links.

The result is a list of unique links of *cmid-pnum* to PIK comprising of all college graduates (at the time of survey) surveyed in the ACS and NSCG. Note that, the sampling design of the NSCG means that NSCG PIKs are a subset of the ACS PIKs (the overlap is the subset of individuals from the ACS selected for follow-up interviews in the NSCG.). But as indicated above, I separate these two groups of individuals so that all individuals who were in the ACS and NSCG are treated as belonging to the NSCG and not the ACS. Thus, ACS individuals are now defined as those were surveyed in the ACS *and not* the NSCG.

3. **random sample** I take a 55% random sample of individuals based on their PIK. This is done for disclosure review purposes. Demographic characteristics for this sample are displayed in Column (1) of Table A.18.

4. **education restrictions** I next drop all individuals who have some form of missing or invalid education. This consists of either a missing or imputed education level or college major. I also restrict individuals to those who are from the 1999-2012 graduating cohorts. I have coverage for all 23 states in the LEHD for years 2000-2014 and keeping individuals from the 1999-2012 cohorts ensures that I have adequate coverage of the early career. Demographic characteristics for this sample are displayed in Column (2) of Table A.18.

5. **employment restrictions** I restrict the sample to only include individuals who were reasonably attached to the labor force in the first year post graduation. Reasonably attached to the labor force is defined as having three-plus quarters of non-zero earnings in the year. This restriction is defined *within* the 23 covered states. I impose this restriction as I am merging a national sample of individuals to a sub-national database of earnings records. Demographic characteristics for this sample are displayed in Column (3) of Table A.18.

   This restriction thus drops individuals who may have annual observations with three plus quarters of non-zero earnings in one of the 23 states in a later portion of their career (e.g. graduated in 1999 and have 3-4 quarters of non-zero earnings in 2001 or any year 2002-2014 but not in 2000).

Table A.17: Sample Selection: Individual- and Observation-Level Restrictions

	ACS-LEHD	NSCG-LEHD	ACS-NSCG
individuals:			
55% sample of 4-year graduates:	ever surveyed in 2009-2019 ACS	ever surveyed in 2010-2019 NSCG	ever surveyed in 2010-2019 NSCG
graduation cohorts:	1999-2012	1999-2012	1999-2015
employment:	3+ qtrs earn>0 in 23 LEHD states in year 1 post-grad	3+ qtrs earn>0 in 23 LEHD states in year 1 post-grad	employed in ACS & first NSCG survey
observations:			
years since grad:	1-15	1-15	1-15
employment:	3+ qtrs earn>0 in covered states	3+ qtrs earn>0 in covered states	non-missing occupation
obs per person:	up to 15	up to 15	2- 5
main use:	earnings growth employer & industry changes	graduate enrollment & degrees	occupation change

Note: Table provides the sample restrictions imposed on individuals and observations for three different analysis datasets.

Table A.18: Individual-Level Sample Restrictions & Person Level Summary Statistics

sample	ACS-LEHD			NSCG-LEHD			ACS-NSCG		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
demographics (%):									
female	.527	.570	.595	.458	.500	.504	.458	.506	.486
black	.060	.060	.048	.092	.103	.096	.092	.103	.100
hispanic	.056	.075	.074	.095	.115	.130	.095	.120	.117
born in a covered state	.358	.373	.710	.323	.333	.644	.323	.333	.343
graduation year (mean)	1987	2005	2005	1994	2006	2006	1994	2007	2007
curriculum-based major groups (%):									
Business & Economics	.195	.190	.198	.097	.092	.096	.097	.092	.091
CS & Engineering	.108	.113	.091	.260	.282	.270	.260	.287	.300
Communications & Marketing	.058	.084	.094	.023	.024	.026	.023	.023	.023
Education	.142	.095	.103	.045	.028	.035	.045	.028	.029
Health	.067	.065	.068	.065	.069	.070	.065	.071	.071
Humanities	.136	.142	.138	.066	.059	.061	.066	.055	.054
Other Majors	.048	.048	.051	.038	.026	.030	.038	.028	.027
Pure Sciences	.115	.116	.097	.219	.208	.183	.219	.209	.200
Social Sciences	.131	.148	.158	.188	.205	.226	.188	.205	.200
general and specific majors (%):									
general	.415	.447	.454	.370	.372	.383	.370	.379	.357
specific	.346	.298	.292	.354	.359	.357	.354	.356	.371
not general or specific	.226	.256	.252	.260	.256	.261	.260	.264	.257
individuals:	5,764,000	1280,000	383,000	96,000	39,000	11,500	96,000	43,500	35,000

Note: Table displays summary statistics for three different samples of individuals: (1) the baseline sample: contains a 55% random sample of all college graduates in the ACS-NSCG master list; (2) the education sample: individuals from (1) restricted to the graduating cohorts 1999-2012 for the ACS-LEHD and NSCG-LEHD and 1999-2015 for the ACS-NSCG; (3) the employment sample: individuals from (2) that had three-quarters of non-zero earnings in the 23 covered LEHD states in the first year post graduation for the ACS-LEHD and NSCG-LEHD, and individuals that were employed in the ACS and their first NSCG survey for the ACS-NSCG. See text for additional details. Results were disclosed by the U.S Census Bureau's Disclosure Review Board with approval number CBDRB-FY21-P2420-R9067. All cell counts are rounded.

A.2.3 Other Variable Definitions

Graduation Year For individuals in the NSCG year of undergraduate degree is observed but for individuals from the ACS year of graduation is unobserved. For ACS individuals, I follow Altonji et al. (2016b) and use year and quarter of birth to impute graduation year as the year in which an individual was aged 22 or 23. Specifically: if birth quarter is 1 or 2, *graduation year* = *birth year* + 22, and if birth quarter is 3 or 4, *graduation year* = *birth year* + 23. Estimates from the public-use NSCG indicate that age 22 is the modal age at graduation (over 30%), followed by age 21 (∼22%) and age 23 (∼17%).

Employer (SEIN) Employer is measured in the LEHD. In the LEHD a worker's earnings observations are tied to an employer's state-level unemployment insurance (UI) account (SEIN), which is the concept of employer used in this paper. In some cases the notion of a SEIN is equivalent to the notion of a firm but in other cases it is smaller. Specifically, a SEIN and firm are equivalent if the firm has only a single establishment nationally. A SEIN is smaller than a firm if a firm has multiple SEINs within a state or a firm has multiple SEINs across states. Defining SEIN as the employer means I will treat two workers whose earnings are reported to under different SEINs, either within state or between states, that belong to the same firm as having different employers. Because the occurrence of multiple SEINs for a firm within a state is relatively rare, I will conceptualize the SEIN as synonymous with a state-level employer, and hereafter refer to a worker's SEIN as the worker's employer.[4]

Employer Characteristics Both employer pay and size are measured at the SEIN-year. Both measures are based only on the earnings records of full-quarter workers. A worker is defined as employed full-quarter in quarter X if they are observed with non-zero earnings in quarter $X - 1$ and $X + 1$. All data come from the Quarterly Workforce Indicators (QWI) SEINUNIT file which is employer-level data based on the aggregated data for *all workers* covered by the LEHD unemployment insurance records in a given time period (e.g. all workers with non-zero earnings).

To construct employer size, start with quarterly measures of full quarter employment at the SEINUIT level, as measure in the restricted-use version of the Quarterly Workforce Indicators (QWI) SEINUNIT file. A state-level employer (SEIN) can be comprised of

[4]The case of multiple SEINS for a firm usually occurs when the firm operates in multiple industries within the state but is not very common. As stated in Sorkin (2018): Firms with multiple SEINs in a state is a rare occurrence: according to personal communication from Henry Hyatt (June 12, 2014): "the employment weighted fraction of firmids with multiple SEINs [state employer identification number] in a given state is about 1.5%, and. . . this fraction is actually lower in some of the larger states.". Although most employers have on establishment (are "single unit"), most employment is with employers who have multiple establishments ("multi-unit").

one-many establishments (SEINUNIT). I convert the data from SEINUNIT to SEIN level by aggregating up multi-unit establishments into one observation per SEIN and summing full quarter employment across all SEINUNIT within a quarter. I next convert the SEIN quarterly measures into SEIN annual measures by taking the average of full quarter employment across all 4 quarters.[5] The result is a measure of the quarterly average of full-quarter employment within a year.

To construct employer pay, I convert SEINUNIT to SEIN observations by summing the quarterly earnings of full quarter workers across all SEINUNIT within a quarter. I next convert the SEIN quarterly measures into SEIN annual measures by taking the average of full-quarter worker earnings across all 4 quarters in the year: I divide quarterly earnings of full quarter workers by full quarter employment in each quarter and take the average of this measure across all 4 quarters. The resulting measure is average quarterly earnings per full-quarter worker.

Employer characteristics are measured in the first year of the employer-employee relationship and are fixed until separation so that changes in employer characteristics only occur with employer changes. Fixing employer attributes at the first year of the employer-employee relationship rules out the case in which a worker began employment for a start-up (small employer size) that was later successful and grew dramatically both in size and average pay.

Industry Industry is the LEHD is attached to an employer identifier. Industries are classified using the employer's two- and four-digit industry North American Industry Classification System (NAICS) codes. These levels of detail correspond to the employer's economic sector and primary business activity, respectively.

Occupation I work with two different aggregates of occupation codes: 83 detailed occupations and 20 broad occupations. To create the codes I first harmonize occupation codes across the American Community Survey (ACS) waves, and then across the ACS and National Survey of College Graduates (NSCG) surveys.

The ACS codes found in IPUMs vary a bit between the 2009, 2010-2011, 2012-2017, 2018-2019 survey waves. I first crosswalk all ACS codes to the 2010 ACS codes. The largest adjustments were between the 2009 ACS codes which are based in the 2000 Census and SOC occupation coding scheme and the 2010 ACS codes which are based in the 2010 Census and

[5]I experiment with 3 different ways of doing this: First, I sum full quarter employment in the year and divide by 4. Second, I sum full quarter employment in the year and divide by the number of quarters with non-zero full quarter employment. Third, I construct a weighted average of full quarter employment with weights equal to each quarter's full quarter employment.

SOC codes. I used the general rules to aggregate codes. If a single 2009 ACS codes mapped to several 2010+ codes, I aggregated the 2010+ codes into a single occupation and assigned it the single 2009 code. If several 2009 codes mapped to a single 2010+ code, I aggregated the 2009 codes. The 2018 and 2019 splits several of the occupation codes found in 2010-2017 ACS and I collapse the 2008 to 2019 codes back into their 2010 codes.

I next aggregated the ACS occupation codes into the NSCG occupation codes based on the names of the occupations. For some large classes of occupations that entail more routine-manual work (e.g. Precision/production occupations and Transportation and Moving Materials occupations) I use David Dorn's crosswalk of the occ2010 to occ1990dd codes. The result is a system of 83 occupations that are harmonized across the ACS and NSCG. Finally, I aggregate the occupation codes into roughly 20 broad occupation codes following Altonji and Zhong (2021). As the NSCG occupation codes vary in their detail across broad groups of occupations, I primarily focus on outcomes based on the broad occupation codes.[6] The <u>broad</u> and narrow occupation codes include:

- <u>Biological scientists</u>: Agricultural and food scientists, Biological scientists, Foresters and conservation scientists, Medical and Life Scientists
- <u>Blue Collar Occupations</u>: Construction and extraction occupations; Installation, maintenance, and repair occupations; Precision/production occupations; Protective services (e.g., fire fighters, police, guards, wardens, park rangers); Transportation and Moving Materials occupations
- <u>Business Related Occupations</u>: Accountants and Auditors and other financial specialists; Actuaries; Business and Financial Operations Occupations (Insurance securities real estate and business services); Personnel training and labor relations specialists
- <u>Clerical Occupations</u>: Accounting clerks and bookkeepers; Other Office and Administrative Support Occupations; Secretaries, receptionists, typists
- <u>Computer Scientists</u>: Computer Network Architects; Computer programmers (business, scientific, process control); Computer system analysts; OTHER computer information

[6]For example Computer Occupations are very disaggregated and include Computer programmers (business, scientific, process control, Computer system analysts, Computer support specialists, Database administrators, Information security analysts, Network and computer systems administrators, other computer information science occupations, Software developers, Web developers and Computer Network Architects. Health Occupations are fairly aggregated with occupations including "Diagnosing/treating practitioners: Physicians, Dentists, Veterinarians, Optometrists, Podiatrists" and "Registered nurses, pharmacists, dieticians, therapists, physician assistants, nurse practitioners". For Business occupations there is a fairly general occupation titled "Business and Financial Operations Occupations (Insurance securities real estate and business services)" but also more specific occupations like "Accountants and Auditors and other financial specialists" and "Chief Executives and Legislators and Top Level Managers".

science occupations; Operations and systems researchers and analysts; Software developers

- Doctors: Diagnosing/treating practitioners (Physicians, Dentists, Veterinarians, Optometrists, Podiatrists)
- Engineers: Aerospace Engineers; Architects; Biomedical and Agricultural Engineers; Chemical Engineers; Civil Engineers; Computer Hardware Engineers; Electrical and Electronics Engineers; Environmental Engineers; Marine Engineers and Naval Architects; Materials Engineers; Mechanical Engineers; Other Engineers; Petroleum, mining and geological engineers; Sales engineers
- Farmers, Foresters and Fisherman: Farmers, Foresters and Fisherman
- Law Related Occupations: Lawyers and judges
- Managers: Chief Executives and Legislators and Top Level Managers; Education administrators (e.g. registrar dean principal); Medical and Health Services Managers; Natural Science Managers; Other Managers
- Marketing: Other Sales, Marketing and Related Occupations; Commodities sales (e.g. machinery, equipment, supplies); Retail Sales (e.g. furnishings, clothing, motor vehicles, cosmetics)
- Math Scientists: Mathematicians and statisticians
- Other Computer occupations: Computer support specialists; Database administrators; Information security analysts; Network and computer systems administrators; Web developers
- Other Health Occupations: Health technologists and technicians (e.g dental hygienists, licensed practical nurses, medical/laboratory technicians); Other Health Occupations; Registered nurses, pharmacists, dietician, therapists, physician assistants, nurse practitioners
- Other Service Occupations: Food preparation and service (e.g., cooks, waitresses, bartenders); Other service occupations, except health
- Other Social Service Occupations: Clergy and religious workers; Counselors; Librarians, Archivists, Curators; Social Workers
- Physical Scientist: Atmospheric and space scientists; Chemists and Materials Scientists; Environmental Scientists and Geoscientists; Other Physical Scientists; Physicists and Astronomists
- Postsecondary Teachers: Postsecondary Teachers
- Primary and Secondary Teachers: Education Workers, Other; Elementary and Middle School Teachers; Preschool and Kindergarten Teachers; Secondary School Teachers; Special Education Teachers

- Social Scientist: Economists; Psychologists; Social Scientists
- Technician: Drafters; Engineering Technologists/Technicians/Surveyors; Life, Physical, and Social Science Technicians; Surveyors, Cartographers, and Photogrammetrists
- Writers and Artists: Arts, Design, Entertainment, Sports, and Media Occupations

Graduate Education Graduate degree types are pre-defined in the ACS and NSCG as follows: Master's (MA) degrees include Master of Science (MS), Master of Arts (MA), and Master of Business Administration (MBD). Doctorate includes Doctor of Philosophy (PhD), Doctor of Science (DSc), and Education Doctorate Degree (EdD). Professional degrees include Juris Doctor (JD), Bachelor of Laws (LLB), Doctor of Medicine (MD), Doctor of Dental Surgery (DDS), Doctor of Veterinary Medicine (DVM).

The NSCG has institution (restricted-use version only), type (BA, masters, professional degree, PhD), major and date for all post-secondary educational degrees ranging from first bachelor's degree, most recent degree and 2nd-5th highest degree. I stack the 2010-2019 wages of the NSCG and collect and sort by award date the degree information for each individual.

As I merge the NSCG graduate degree information onto the LEHD, and the 2014 LEHD snapshot only has data through year 2014, I only pay attention to degrees with enrollment or attainment periods that would fall into the 2000-2014 LEHD analysis window. I use the degree information to code each individual's annual observations from 200-2014 as either pre-enrollment, during enrollment and post-graduate degree attainment. While the NSCG does have enrollment variables, the timing of the NSCG (2010-2019) is mostly outside of the 2000-2014 LEHD data window and so I instead follow Altonji & Zhong (2021) and impute enrollment dates by assuming 2,3 and 5 years of enrollment for master's, professional degrees and bachelor's degrees.

For workers who have one graduate degree it is simple: if graduated with a master's in May 2010 then I code your attainment as 2011+ and start the slope in 2011. Enrollment dummy is equal to one in 2009 and 2010. For a two-year degree with a May 2010 graduation day that implies an August 2008 start period. Since the vast majority of 2008 is not enrolled, I decide to treat that year as pre-enrollment, and treat 2011 as first year post graduation to be consistent with the undergraduate years since graduation timing. I treat professional degrees and PhD in a similar fashion but vary the length of enrollment period (3 years for professional degree and 5 years for PhD). If a worker had two graduate degrees and degree dates imply a continuous period of enrollment, I only code the worker as having attained a graduate degree upon receipt of the second degree (example, receiving an embedded master's with a PhD or attaining a master's and directing enrolling in a PhD).

A.2.4 Comparing outcomes in the ACS-LEHD and NSCG-LEHD

For the vast majority of the empirical analysis I utilize the ACS-LEHD analysis sample, but for some supplementary analyses I use the unweighted and weighted NSCG-LEHD sample. In this section, I discuss differences in the outcome trajectories across three samples: (1) ACS-LEHD, (2) the unweighted NCSG-LEHD and (3) the weighted NSCG-LEHD.

For each labor market outcome, I compare the implied trajectories with years since graduation across the three samples and two difference specifications. The first specification parameterizes years since graduation using a two-year bins:

$$Y_{it} = \beta_0 + \sum_g \beta_g g_{it} + \sum_m \phi_m(\text{major group}_{im}) + \sum_m \sum_g \beta_{m,g}(g_{it} \cdot \text{major group}_{im}) \\ + \text{controls} + \epsilon_{it} \tag{A.1}$$

so that g_{it} is equal to one if year t is either year since graduation in the bin g. In the second function, years since graduation are indexed using a quadratic function:

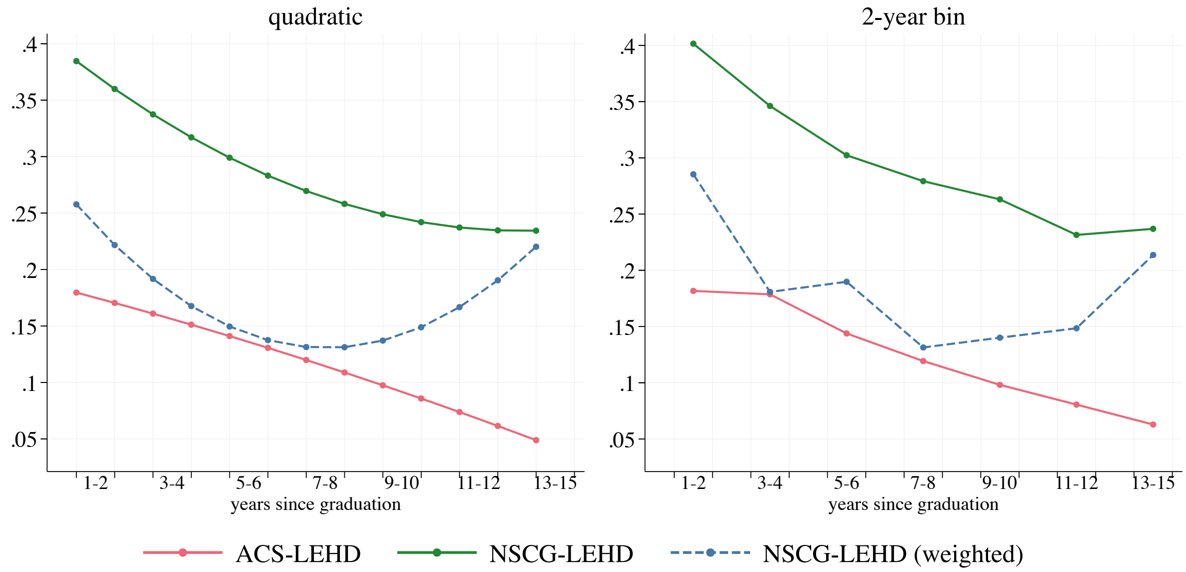
$$Y_{imt} = \beta_0 + \beta_g g_{it} + \beta_{g2}(g_{it})^2 + \sum_{m \in M} \phi_m(\text{major group}_{im}) + \sum_{m \in M} \beta_{m,g}(g_{it} \cdot \text{major group}_{im}) \\ + \sum_{m \in M} \beta_{m,g2}((g_{it})^2 \cdot \text{major group}_{im}) + \text{controls} + \epsilon_{it} \tag{A.2}$$

Log Earnings Gap between Specific and General Majors Figure A.11a displays the evolution of the specific major's earnings premium for the three datasets and the two regression specifications. In the left panel, which displays the results from the quadratic specification, the log earnings premium of specific majors in the ACS-LEHD decreases steadily from .18 initially to .05 13-15 years post graduation. For the unweighted NSCG-LEHD sample, the decrease is also monotonic but the earnings gap is of a larger magnitude: it starts at around .4 and decreases to around .25. The magnitude of the weighted NSCG-LEHD earnings gap falls between that of the two samples but follows a u-shape: initially the log earnings gap is .25 and decreases steadily to .15 7-8 years post graduation, after which point it increases again to .25. An inspection of the two-year bin model shows that this is entirely driven by an increase in the earnings gap in the period of 13-15 years post graduation: the earnings gap in years 9-10 and 11-12 post graduation is around .15 but in 13-15 years post graduation it jumps to over .2. Figure A.11b shows that this is due to a steep drop in the earnings growth of general majors from 11-12 to 13-15 years post graduation, but no equivalent drop among specific majors.

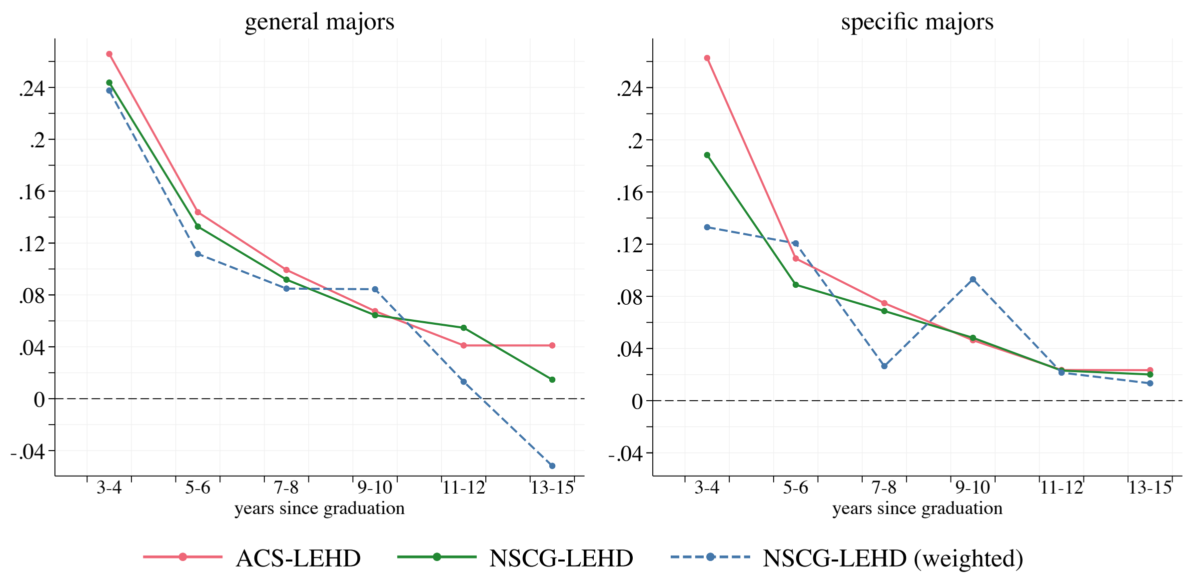
Employer and Industry Changing Figure A.12a displays the gap between general and specific majors in the frequency of employer changing for the three datasets and the two regression specifications. In the quadratic specification, and in all three datasets, general majors change employers more often than specific majors throughout the first 15 years of the career. In the ACS-LEHD the percentage point differential decreases steadily. In both the unweighted and weighted NSCG-LEHD the differential follows a u-shape pattern, because results from the 2-year bin specification (right panel) show that the gap between general and specific majors in job changes increases substantially 13-15 years post graduation. Figure A.12b shows that the same general pattern of differences across datasets and regression specifications holds for industry changes as does for employer changes.

Figure A.11: Comparing Earnings across Datasets: ACS-LEHD and NSCG-LEHD

(a) Specific Majors' Earnings Premium



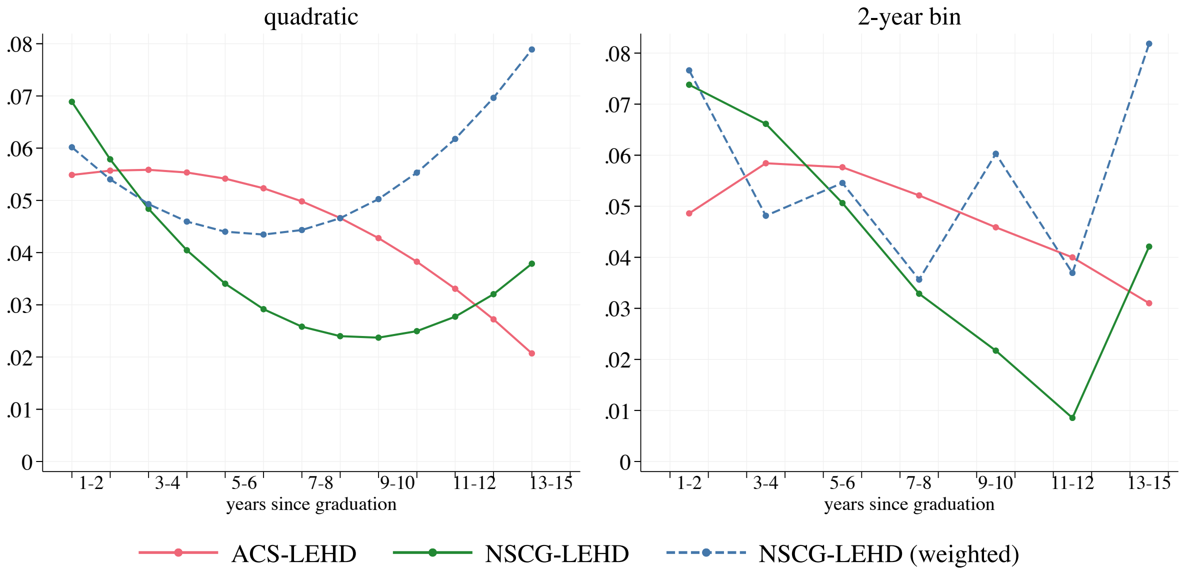
(b) Log Earnings Growth from Last Period



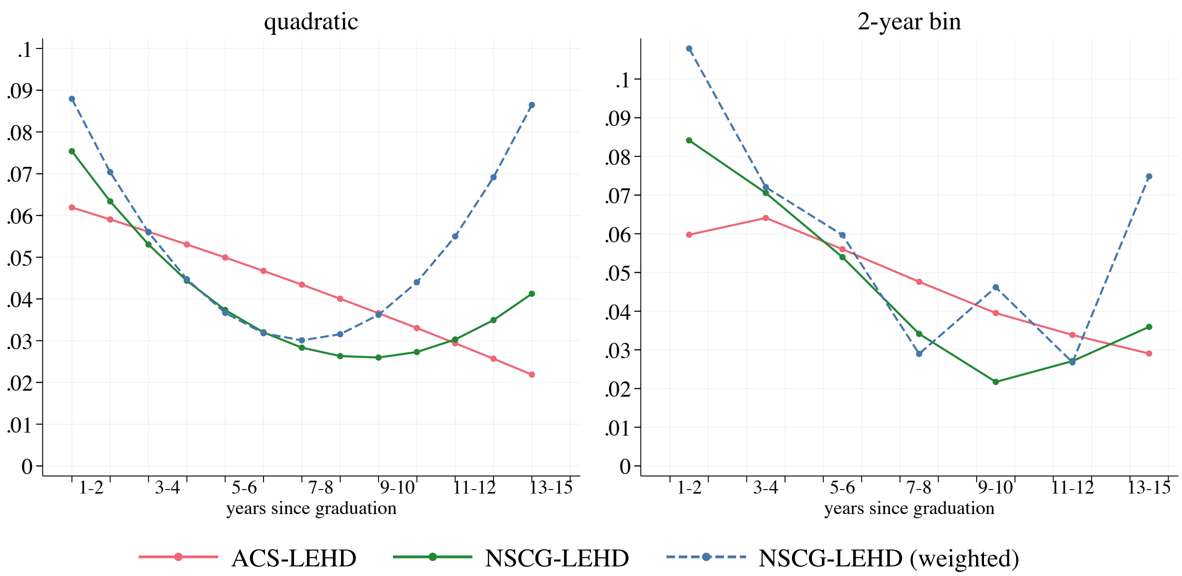
Note: Figure displays estimates the specific majors' earnings premium and the log earnings growth from the previous period for general and specific majors for two difference specifications: a specification with years since graduation indexed by two-year bins (Equation (A.1)) and by a quadratic function (Equation (A.2)). The results are plotted for three different analysis samples: the ACS-LEHD, the unweighted NSCG-LEHD and the weighted NSCG-LEHD. Data source is the ACS-LEHD. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

Figure A.12: Comparing Job Changing across Datasets: ACS-LEHD and NSCG-LEHD

(a) Employer Changing Gap: General-Specific Majors



(b) Industry Changing Gap: General-Specific Majors



Note: Figure displays estimates of the difference between general and specific majors in employer and industry changes for two difference specifications: a specification with years since graduation indexed by two-year bins (Equation (A.1)) and by a quadratic function (Equation (A.2)). The results are plotted for three different analysis samples: the ACS-LEHD, the unweighted NSCG-LEHD and the weighted NSCG-LEHD. Data source is the ACS-LEHD. Results were disclosed by the U.S Census Bureau's Disclosure Review Board. All point estimates are rounded.

APPENDIX B

Appendix to Chapter 2

B.1 Defining Major Categories

To aggregate the almost 400 four-digit majors of the CIP taxonomy into a smaller set of 70 aggregated categories (hereafter referred to as *final major*), we start with the CIP's aggregation of four-digit majors (cip4) into 49 two-digit major codes (cip2). We omit from our categorization 14 two-digit categories that are traditionally sub-baccalaureate or remedial programs (Interpersonal and Social Skills (cip2=35), Basic Skills and Developmental/Remedial Education (32), Citizenship Activities (33), Health-Related Knowledge and Skills (34), Personal Awareness and Self-Improvement (37), High School & Secondary Diplomas and Certificates (53)); that are predominantly post-baccalaureate or graduate programs (Residency Programs (60)); that are predominantly trade-specific and usually sub-BA (Science Technologies/Technician (41), Construction Trades (46), Mechanic and Repair Technologies/Technicians (47), Precision Production (48), and Transportation and Materials Moving (49)); or that operate in separate or specific labor markets (Military Science, Leadership, and Operational Art (28) and Military Technologies and Applied Sciences (29)). Together these categories comprise less than 1% of all degrees granted by four-year postsecondary institutions over the 2010-2017 period and appear on less than 0.1% of job postings in our analytic sample. For similar reasons we also omit particular four-digit majors (not already in omitted two-digit categories) that are primarily sub-baccalaureate or graduate programs, including Funeral Service and Mortuary Science (1203), Cosmetology and Related Personal Grooming Services (1204), Medical Clinical Sciences/Graduate Medical Studies (5114), Chiropractic (5101), and Dentistry (5104).

For the remaining two-digit categories, we calculate the total number of job postings shared among the four-digit majors contained in the two-digit category. Two-digit major categories that have few postings (less than 0.1%, or about 22,000 unique postings in our sample) are aggregated together as described below. For the large two-digit major categories we make a few general adjustments. First, we pull out some four-digit majors that are particularly large in terms of job postings. For example, in the two-digit category Architecture and Related Services (cip2=04), the four-digit major Architecture (cip4=0402) accounts for more than half of postings and degrees granted for the two-digit category. We thus split the two-digit category into the two *final major* groupings of (1) Architecture and (2) Urban and Regional Planning and Design. For the two-digit group Social Sciences (cip2=45), we disaggregate the four-digit majors of Sociology (cip4=4511), Economics (cip4=4506), and Geography (cip4=4507), all of which have large numbers of job postings and four-year degrees granted during 2010-2017, into three separate *final majors*, combine International Relations and National Security Studies (cip4=4509) and Political Science and Government (cip4=4510) into another *final major* and aggregate most of the remaining four-digit majors into a *final major* called Other Social Sciences. As a final example, the 15 four-digit majors in the broad category of Education are grouped into three *final major* categories: (1) Special Education and Teaching, (2) Teacher Education, and (3) Other Education.

In some cases, pulling an individual four-digit major out of a two-digit category would result in an aggregation of the other remaining four-digit majors with a relatively small number of job postings. In these cases, we do not disaggregate the two-digit category; instead the two-digit category remains a *final major* category. For example, in the broad category of Family and Consumer Sciences & Human Sciences (19), the four-digit major Human Development, Family Studies, and Related Services (1907) constitutes over 86% of postings for the two-digit category, and the entire two-digit family becomes *final major* Family and Consumer Science. In other cases, although individual four-digit majors have both a large number of postings and degrees granted, the four-digit majors are commonly co-listed together on job postings. We aggregate these four-digit majors together into a *final major*. For example, within the two- digit category of Computer and Information Sciences and Support Services (11) the three most frequently occurring four-digit majors of Computer and Information Science, general (1101), Computer Science (1107), and Information Sciences/Studies (1104) are often listed on job postings together.

Finally, there are a few particular two-digit major categories that we split into more narrow *final major* categories, based on similarity of content or labor market outcomes. For example, in the broad category of Engineering there are over 39 four-digit majors which we aggregate into 10 *final major* categories including Mechanical Engineering, Computer

Engineering, Electrical Engineering, and Civil Engineering. The 35 four-digit majors within the two-digit category Health Professions and Related Programs are aggregated into *final major* categories including Allied Health, Mental and Social Health Services, and Nursing.

We next deal with two-digit major categories that have few job postings, including Area, Ethnic, Cultural and Gender Studies (cip2=05), Communications Technologies/Technicians and Support Services (cip2=10), English Language and Literature/Letters (cip2=23), Liberal Arts and Sciences, General Studies Humanities (cip2=24), History (cip2=54) and Multi/Interdisciplinary Studies (cip2=30). To find the best fitting *final major* categories for each of these, we calculate the skill distance between the group and other four-digit majors. Generally, we use this method to find for each four-digit major the closest other four-digit majors, and assign it to the same *final major* category. Specifically, for each major we calculate the proportion of category postings for each of 8 skill composites ([# of ads with skill=s & majorcat=c]/[# of ads with majorcat=c]) on a sub-sample of our data. We then use the proportions to calculate a measure of cosine similarity: $\frac{\sum_s^{s=8}(a_i \times b_i)}{\sqrt{\sum_s^{s=8}(a_i)^2}\sqrt{\sum_s^{s=8}(b_i)^2}}$ where a and b are two different majors and a_i and b_i are the share of major $a's$ and major $b's$ postings that demand skill composite i, respectively. Finally, for a given major we sort other majors based on how similar skill demand is according to the cosine similarity measure. Using this method, we decided to combine the three two-digit majors of English, Liberal Arts and Humanities, and History into one *final major*, and the two-digit category Area Studies into the *final major* Other Social Sciences. We also used this method to find the most similar four-digit major for each of the majors in the fairly heterogeneous two-digit group of Multi/interdisciplinary Studies. As a result, we aggregated Systems Science and Theory (3006) into Management Information Systems and Science (5298), Museology/Museum Studies (3014) into Library Science (2500), and Behavioral Sciences (3017) into Psychology (4200).

B.2 Constructing Skill Composites

We initially followed the keyword approach of Deming and Kahn (2018) to allocate individual skills to skill composites. Our decision to reallocate individual skills to composites stemmed from three observations about the skill-to-composite mappings resulting from the keyword approach.

First, some of the most frequently listed skills did not fall into any skill composite. Examples include planning (20% of postings), organizational skills (16%), detail-oriented (12%), scheduling (12%), building effective relationships (11%), creativity (10%), troubleshooting (6%) and multi-tasking (8%).

Second, our use of the keyword approach meant that some skills were misclassified. The most prominent example is the case of using the keyword "management" to allocate skills to the skill composite "people management". The term "management" captures a wide variety of general management activities that do not specifically pertain to HR or personnel, including account management, pain management, operations management, case management, and management consulting. Another example was character (organizational) skills, which was initially defined as keywords "organized, detail-oriented, multitasking, time management, meeting deadlines, energetic" and as a result missed the very common variant skills of "multi-tasking" and "organizational skills".

Third, the ill-fitting mapping of skills to composites occurred for some of the most-frequent skills. In the case of relatively rare skills, misclassification of individual skills can be viewed as a form of measurement error that should not have a large impact on empirical results. However, since some individual skills are sufficiently common and get assigned to composites that seem incorrect a priori, we believe misclassification may bias the interpretation of a given skill composite. Thus, we focus on reallocating the individual skills that appear with the highest frequency.

We use the following procedure to map the 1,000 most frequent individual skills listed on job postings that demand a bachelor's degree to 11 skill composite categories. (The 1,000th most frequent skill appears on 0.2% of job postings that demand 16 years of education.) First, for each individual skill, two different individuals on the research team independently assigned the skill to one of the 11 categories according to the definition of the skill categories shown below. In roughly 40% of cases, two individuals assigned an individual skill to different skill composites. For the 10 most frequent skills in which individual coding to composites differed, we discussed as a group which skill composite would be most fitting. We then refined our skill composition definitions, and pairs of individuals revisited and resolved cases in which a single skill was assigned to multiple skill composites. After this step there remained roughly 50 individual skills that pairs of reviewers still believed could fit into multiple categories. We allocated these skills to a single skill composite by consulting the occupation distribution of ads listing the skill. Table B.1 displays the final number of individual skills, and the three most frequent skills, allocated to each skill composite. Appendix Table B.4 shows the assigned skill composite for the 40 most frequently listed skills.

- **Social**: Communicating, persuading, or negotiating with others, which involves adept presentation or exchange of information and perspectives as well as the capacity to accurately infer the motivations of others.
- **People Management**: Supervising, motivating, or directing people internal to the business toward defined goals.

- **Cognitive**: Applying analytic, logical, quantitative or qualitative reasoning, evaluation, or critical thinking to understand patterns and solve problems.
- **Writing**: Composing, drafting, and editing of books, papers, reports, releases, scripts and other text-based documents; excludes underwriting (which is cognitive).
- **Customer Service/Client management**: Attracting, soliciting, maintaining, and retaining clients and customers; most forms of sales fall here if there is a personal contact (sales engineering or analysis is cognitive).
- **Organization**: Organizing, planning, managing, and expediting meetings, conferences, events, and other time-sensitive activities; but not logistics or supply chains (which are project management); ability to balance and prioritize among competing demands, apportion work, and meet deadlines.
- **Computer**: General computer tasks and knowledge, including MS Office and related frontline computer support; excludes computer engineering, hardware, design, and other specialized tasks.
- **Software**: Use or design of any specialized software, as well as any computer hardware design and engineering, and computer security or network management.
- **Financial**: Preparing or auditing payroll, budgets, accounting or tax documents, and financial reports and statements; excludes financial trading (social), financial engineering, or quantitative financial analysis (both cognitive)–the distinction is that the financial composite captures highly prescribed and rules-based activities that are often ancillary to main activities (unless the main activity is auditing/accounting).
- **Project Management**: Orchestrating, overseeing, or directing programs, projects, processes, and operations–the distinction with people and client management is that the emphasis here is not on people, but rather on the substance of the plans and activities executed by people.
- **Other**: Highly discipline-specific skills (often in health) or physical skills that do not readily generalize to other tasks.

B.3 Hand-Coded vs. Keyword Skill Composites

Our preferred approach to classifying skills was to assign by hand the 1,000 most frequent skills, as described above. This Appendix describes the sensitivity of our approach to the alternative of using the keywords displayed in Table B.1 to identify skill composites.

B.3.1 Coverage

For all composites except software and people management, the share of ads assigned to the composite increases with our approach. About 1 in 500 postings do not list any of our 11 composites; this figure was closer to 1 in 25 based on the keyword approach, which covered only 8 composites. Notably, the keyword approach captured only 400 of the 1000 most frequent skills, while our preferred approach classifies all 1000. Preferred composites are now mutually exclusive: under the keyword approach, about 200 individual skills fell into more than one composite (70% of these involve software, and 30% involve customer service, people management, and cognitive).

The composites under our preferred approach capture a different number of individual, detailed skills than does the keyword approach. Under the latter system, for example, character (organization) contained only three detailed skills: "time management," "meeting deadlines," and "energetic." Our preferred method also captures "multi-tasking," "prioritizing tasks," and "organizational skills." This change means that some of the most common skills are now classified as "organizational skills," as shown in Table B.1 below.

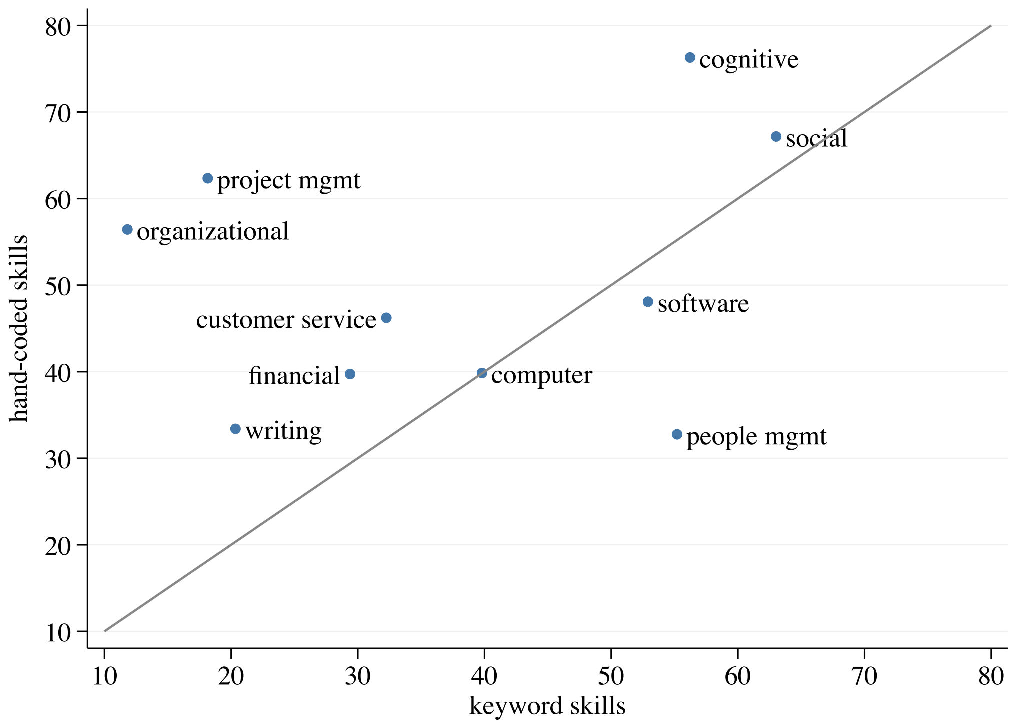
Table B.1: Hand-coded vs Keyword Skill Composites

		Hand-coded		Keyword	
	Skill Composite	N skills in 1000 most frequent	N skills among all skills	N skills in 1000 most frequent	N skills among all skills
1	social	56	56	15	78
2	people mgmt	43	43	85	476
3	cognitive	168	168	46	431
4	writing	20	20	8	50
5	customer service	110	110	56	372
6	organizational	37	37	3	3
7	computer	22	22	12	64
8	software	233	233	175	1703
9	financial	84	84	19	113
10	project mgmt	111	111	1	476
11	other	116	116	–	–
	unclassified	0	14,260	602	12,081

B.3.2 Share of Ads in Each Composite

Figure B.1 below compares the share of unique ads that contain each skill composite across the two different classification approaches.

Figure B.1: Keyword (Old) vs Hand-coded (New) Skill Composites - % of Unique Ads



Note: Figure plots the percent of unique job postings that demand each skill composite. "Keyword" skills refer to the Deming and Kahn (2018) versions of the skill composites and "hand-coded" refers to the versions from Hemelt et al. (2021).

B.3.3 Characterization of Major Skill Concentration

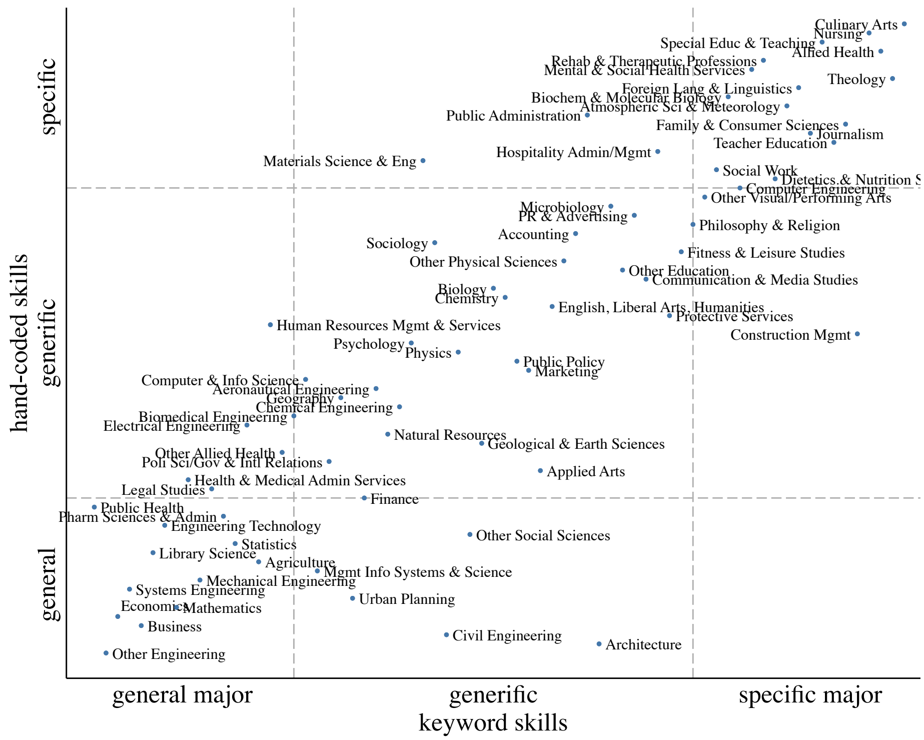
Figure B.2 compares our classification of major skill concentration between the two methods for classifying skills into composites. Figure B.2a compares rank correlation between the two measures; 52 of 70 final majors stay in the same broad category (general, generific, specific) when shifting from the keyword approach to our preferred hand-coding approach.[1] Specifically, 12 majors are "general" (bottom left grouping) under both schemes, 24 stay "generific" (central grouping), and 16 stay "specific" (top right grouping). Nine majors become more specific when switching from the keyword to hand-coding method: for example,

[1]The "general" category includes majors ranked 1 through 18 based on location quotient (LQ) similarity, "generific" includes those ranked 19 through 51, and "specific" includes those ranked 52 through 70. These roughly correspond to the top quartile, middle half, and bottom quartile of majors.

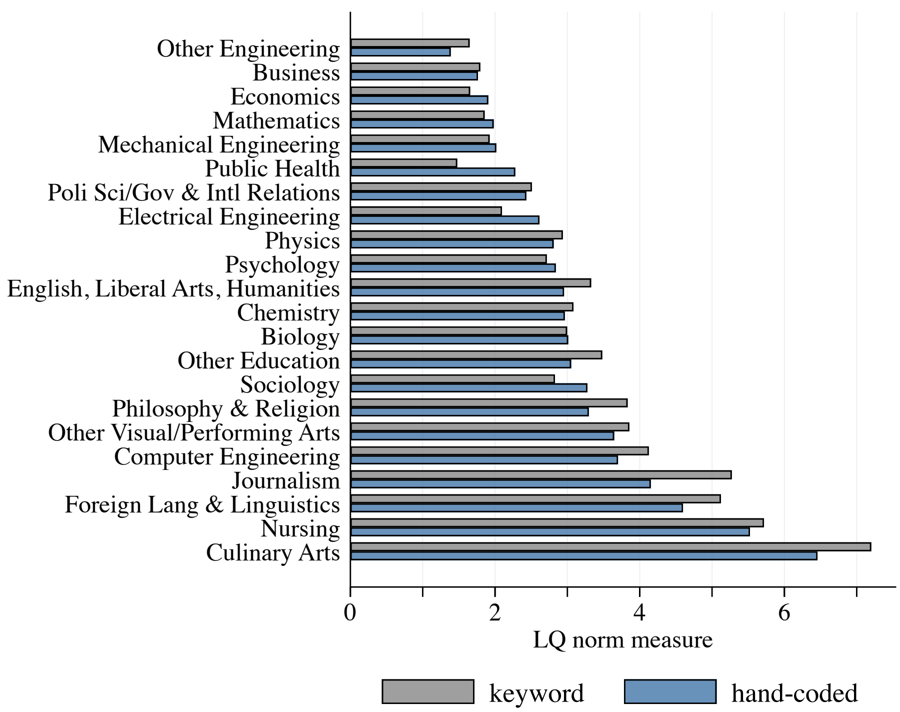
Biomedical Engineering and Legal, which move from “general” to “generific”, and Material Sciences & Engineering and Public Administration, which move from “generific” to “specific.” The last set of nine majors becomes more general, including Philosophy and Other Visual & Performing Arts, which move from “specific” to “generific,” and Architecture and Other Social Sciences, which move from “generific” to “general.” Figure B.2b shows the specificity of selected majors under the two categorization systems in bar chart form.

Figure B.2: Skill Specificity of Majors Using Different Methods to Classify Skills

(a) Rank Correlation



(b) Measure of Skill Specificity



Note: Figure plots the percent of unique job postings that demand each skill composite. "Keyword" skills refer to the Deming and Kahn (2018) versions of the skill composites and "hand-coded" refers to the versions from Hemelt et al. (2021).

B.4 Replication of Deming and Kahn (2018)

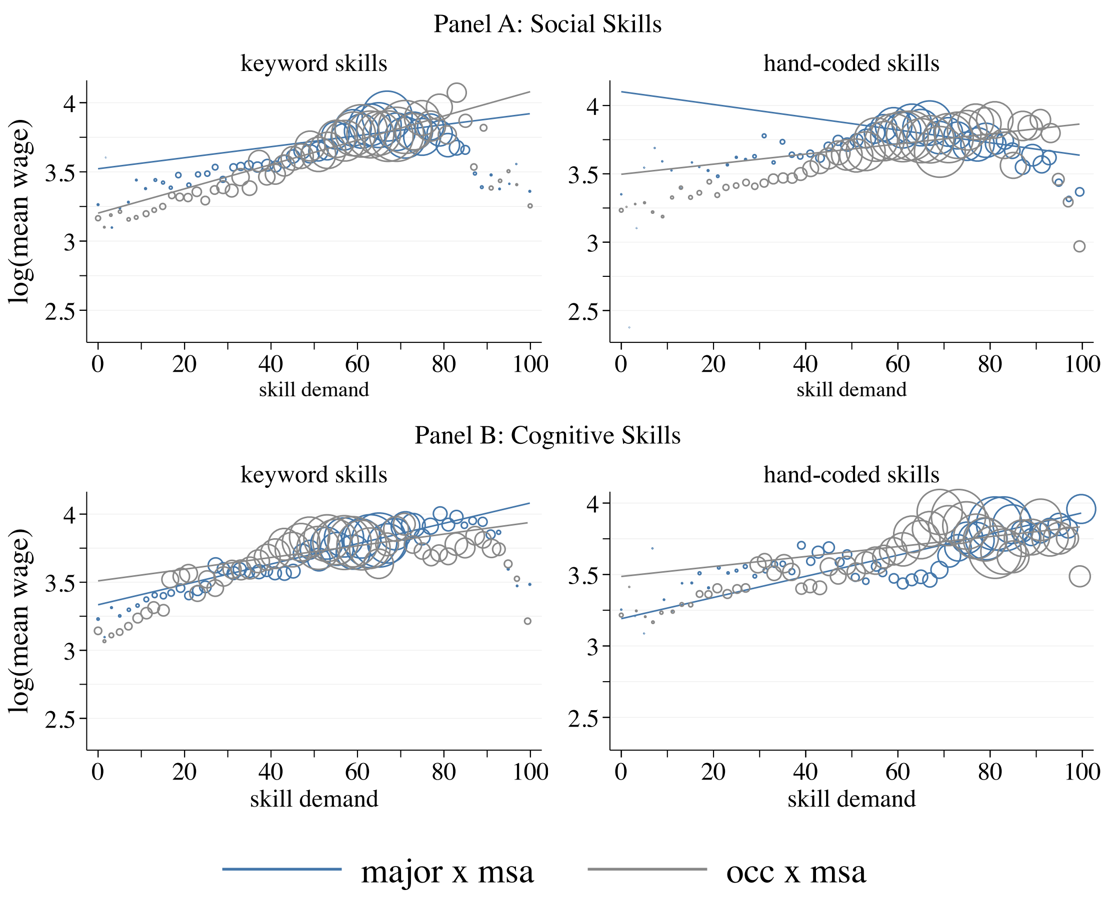
In order to better understand how our findings compare to those of Deming and Kahn (2018) (DK), we attempt to replicate and extend their main cell-level analysis. DK regress log mean wages in a MSA-occupation cell on shares of job ads seeking cognitive skills, social skills, and their interaction. They control for average years of education and experience, the share of ads with each of eight other job skills, and an increasingly rich set of job characteristics, such as MSA and six-digit occupation fixed effects. Their main finding is that cognitive and social skill requirements are positively correlated with wages, both with and without rich controls. Their specification with the most complete set of controls finds that a 10 percentage point increase in the share of ads requiring cognitive (social) skills is associated with 0.8% (0.5%) higher wages. They conclude that skill requirements in local labor markets influence local wages even within narrowly defined occupations.

This conclusion contrasts with our finding of minimal association between skill requirements and major premia after netting out MSA and major fixed effects. These differences could stem from several factors, including the range of education levels considered in job postings, the years of job ad data included, the way in which skill composites are constructed, the vintage of the BGT data, the weighting scheme, and the type of aggregation (occupation vs. major). To assess the importance of these factors we replicate some of the main results found in DK's Table 3. Specifically, we follow DK and construct the log of average hourly earnings in MSA-by-six-digit-occupation cells using Occupation Employment Statistics (OES) data from 2012-2015. We then reconstruct the sample of job postings to match DK's by including job postings irrespective of the required education level. We collapse the data to MSA-by-occupation cells rather than MSA-by-major cells. Finally, we measure skill demand using both versions of the composites: the keyword approach used by DK and our hand-coded composites. Table B.11 presents our replication results.

We are able to replicate the main, fully controlled estimates reasonably well (column 1). Differences in the sample (column 3 vs. 1) have little influence on the estimates; however, the method for classifying skills does. Social skill requirements classified using the keyword approach have a positive association with earnings, but the association is zero or even negative when skills are hand-classified (columns 2 and 4). The final four columns report results for our sample, which aggregates ads into MSA-by-major cells and includes a full set of MSA and major fixed effects. We assess the importance of weighting and the classification method. The final column is quite similar to our preferred estimates in Table 2.5. The classification method and weighting scheme both matter. Estimates are closer to zero when we weight by incumbent workers (as measured in the ACS) rather than by job ads.

We were less successful in replicating the estimates from more parsimonious specifications in column 1 of DK's Table 3. However, in Table B.12 we present raw cell-level correlations between social and cognitive skill requirements and wages, where cells are constructed either by MSA-occupation or MSA-major. Cognitive skill requirements are consistently positively associated with cell-level wages regardless of aggregation process, weighting, or classification method. However, the patterns for social skills are not robust–the keyword approach generates positive associations with wages, but the hand-coding approach generates weaker or even negative associations. These patterns also appear in Figure B.3, which presents scatter plots of cell-level skill demand and wages. This analysis reinforces our conclusion that the skill classification process, weighting scheme, and the manner in which ads are aggregated all contribute to differences between our results and those of DK. Further, the association between social skills and wages is much more sensitive to these choices than is the relationship between cognitive skills and wages.

Figure B.3: Correlation between Cell-level Skill Demand and Wages



Note: Figure plots the binned averages of log(mean wage) across MSA-major (blue) and MSA-occupation (gray) cells. The cells for each category are divided into 50 bins, shown along the x-axis, based on the share of job postings in the cell that specify the indicated skill; each bin is thus two percentiles wide. The y-axis plots the average of log(mean wage) for all cells in the bin. A cell's log(mean wage) is the log of the average wage across individuals employed in the MSA-major or MSA-occupation, as captured in the ACS. Circles are sized based on the total number of job postings in the bin. "Keyword" skills refer to the skill composites from Deming and Kahn (2018) and "hand-coded" refers to the procedure described in the text.

B.5 Appendix Tables & Figures

Table B.2: Explained Variation in Whether a Job Posting Lists at least One College Major

	(1)	(2)	(3)	(4)	(5)	(6)
Model SS	10928.4	12747.2	13138.9	21199.4	22288.4	25544.6
Residual SS	75920.3	74101.5	73709.7	65649.2	64560.2	61304.1
Total SS	86848.7	86848.7	86848.7	86848.7	86848.7	86848.7
R-squared	0.1258	0.1468	0.1513	0.2441	0.2566	0.2941
Adjusted R-squared	0.1218	0.1428	0.1473	0.2395	0.251	0.2722
Baseline variables	x	x	x	x	x	x
f(n skills)	–	x	x	x	x	x
Skill composites	–	–	x	x	x	x
500 most frequent skills	–	–	–	x	–	–
1000 most frequent skills	–	–	–	–	x	–
9000 most frequent skills	–	–	–	–	–	x
Number of variables	1611	1614	1625	2125	2624	10574
Number of skill dummies	0	0	0	500	999	8949
Observations	350,233	350,233	350,233	350,233	350,233	350,233

Note: Table presents regression estimates from six separate regressions. The dependent variable is an indicator for whether or not a job posting lists at least one college major. For computational expedience, we use a 1% sample of all Burning Glass data job postings that require a bachelor's degree. The baseline variables include 941 metro- and micro- statistical region fixed effects, 99 year-by-month fixed effects, 504 six-digit occupation codes and 96 two-digit industry codes. F(skills) is a cubic in the number of skills per job posting.

Table B.3: Has-Major F-Test

	Number of variables	Partial SS	F-test
Occupation (soc6)	482	7769.44	134.88***
Industry (naics2)	96	971.11	47.46***
Internship	1	84.13	386.52**
Year-by-month FEs	99	44.45	2.05***
Metro- / micro- statistical area	932	494.87	7.51***

Note: The table presents F-tests on blocks of covariates from a model in which an indicator for whether or not a job posting lists at least one college major is regressed on 941 metro- and micro- statistical region fixed effects, 99 year-by-month fixed effects, 504 six-digit occupation codes and 96 two-digit industry codes. Some fixed effects are omitted due to singleton observations. The sample is a 1% sample of all postings that require a bachelor's degree. Partial SS is the partial sum of squares from an ANOVA analysis of the baseline model and indicates the magnitude by which total sum of squares would decrease in a model that excludes the block of covariates. Authors' analysis of BGT job postings data.

Table B.4: Categorization of 40 Most Frequently Listed Skills

N	Individual Skill	Composite	N	Individual Skill	Composite
1	Communication Skills	social	21	Microsoft Word	computer
2	Planning	organization	22	Troubleshooting	cognitive
3	Microsoft Excel	computer	23	Accounting	financial
4	Teamwork / Collaboration	social	24	Multi-Tasking	organization
5	Problem Solving	cognitive	25	SQL	software
6	Organizational Skills	organization	26	Staff Management	people mgmt
7	Microsoft Office	computer	27	Customer Contact	customer service
8	Budgeting	financial	28	Presentation Skills	social
9	Research	cognitive	29	Quality Assurance and Control	project mgmt
10	Writing	writing	30	Time Management	organization
11	Project Management	project mgmt	31	Verbal / Oral Communication	social
12	Customer Service	customer service	32	Leadership	people mgmt
13	Sales	customer service	33	Software Development	software
14	Detail-Oriented	organization	34	Analytical Skills	cognitive
15	Written Communication	writing	35	Business Development	customer service
16	Scheduling	organization	36	Physical Abilities	other
17	Computer Literacy	computer	37	English	social
18	Building Effective Relationships	social	38	Patient Care	customer service
19	Creativity	cognitive	39	Oracle	software
20	Microsoft Powerpoint	computer	40	Teaching	social

Note: Table lists the 40 most commonly listed skills in the Burning Glass data and authors' skill composite.

Table B.5: Complete List of Major Aggregates

Code	Name
0100	Agriculture
0300	Natural Resources
0402	Architecture
0499	Urban and Regional Planning and Design
0904	Journalism
0909	Public Relations, Advertising, and Applied Communication
0999	Communication and Media Studies
1100	Computer and Information Science
1205	Culinary Arts
1310	Special Education and Teaching
1398	Teacher Education
1399	Other Education
1402	Aeronautical Engineering
1405	Biomedical Engineering
1407	Chemical Engineering
1408	Civil Engineering
1409	Computer Engineering
1410	Electrical, Electronics and Communications Engineering
1419	Mechanical Engineering
1497	Systems, Industrial, Manufacturing, and Operations Engineering
1499	Other Engineering
1500	Engineering technology
1600	Foreign Language and Linguistics
1900	Family and Consumer Sciences
2200	Legal Studies
2499	English, Liberal Arts, Humanities
2500	Library Science
2602	Biochemistry, Biophysics and Molecular Biology
2605	Microbiology
2699	Biology
2705	Statistics
2799	Mathematics
3100	Fitness, Recreation and Leisure Studies
3800	Philosophy and Religion
3900	Theology
4004	Atmospheric Sciences and Meteorology
4005	Chemistry
4006	Geological and Earth Sciences/Geosciences
4008	Physics
4019	Materials Science and Engineering
4099	Other Physical Sciences
4200	Psychology
4300	Protective Services
4404	Public Administration
4405	Public Policy
4407	Social Work
4506	Economics
4507	Geography
4510	Political Science, Government, and International Relations
4511	Sociology
4599	Other Social Sciences
5098	Design, Photography, Video, and Applied Arts
5099	Other Visual/Performing Arts
5107	Health and Medical Administrative Services
5109	Allied Health Diagnostic, Intervention, and Treatment Professions
5115	Mental and Social Health Services and Allied Professions
5120	Pharmacy, Pharmaceutical Sciences, and Administration
5122	Public Health
5123	Rehabilitation and Therapeutic Professions
5131	Dietetics and Clinical Nutrition Services
5138	Registered Nursing, Nursing Administration, Nursing Research and Clinical Nursing
5199	Allied Health
5203	Accounting and Related Services
5208	Finance and Financial Management Services
5209	Hospitality Administration/Management
5210	Human Resources Management and Services
5214	Marketing
5220	Construction Management
5298	Management Information Systems and Science
5299	Business, general

Table B.6: Top Skills Associated with Three Majors

Economics		Teacher Education		Journalism	
Skill	%	Skill	%	Skill	%
Economics	.99	Early Childhood Educ	.68	Journalism	1.00
Communication Skills	.52	Teaching	.62	Writing	.67
Microsoft Excel	.46	Child Development	.46	Editing	.62
Research	.33	Child Care	.43	Communication Skills	.51
Planning	.25	Organizational Skills	.31	Creativity	.41
Problem Solving	.25	Communication Skills	.28	Social Media	.39
Accounting	.24	Lesson Planning	.26	Research	.32
Teamwork, Collaboration	.24	Health Education	.19	Teamwork,Collaboration	.30
Microsoft Powerpoint	.21	Planning	.18	Organizational Skills	.26
Budgeting	.21	Teamwork,Collaboration	.17	Detail-Oriented	.25

Note: Table presents 10 of the most commonly listed skills on the postings for three different majors in the Burning Glass data. The number of postings that list Economics majors is 607,518, that list Education majors is 97,314, and that list Journalism majors is 211,471.

Table B.7: Share of Ads for Each Major Indicating Demand for Each Skill Composite

Major	Cognitive	Social	Project Mgmt	Organizational	Software	Customer Service	Computer Service	Financial	Writing	People Mgmt	Communication	Other Skills (> top 1000)	Other Skills (< top 1000)
Agriculture	80	66	64	58	13	43	48	47	26	37	44	58	79
Natural Resources	91	64	60	66	21	29	42	42	52	37	45	59	93
Architecture	75	66	69	73	62	30	45	46	34	30	42	34	88
Urban Planning	81	68	63	87	38	32	47	47	48	31	43	46	100
Journalism	76	90	44	74	34	40	47	21	100	26	51	35	85
PR & Advertising	80	93	56	76	31	65	52	34	70	30	56	32	85
Communication & Media Studies	77	90	58	73	37	60	52	31	70	32	56	31	82
Computer & Info Science	82	65	70	50	94	39	27	19	36	29	47	25	84
Culinary Arts	60	43	34	65	1	48	56	75	12	68	20	93	40
Special Educ & Teaching	66	89	20	47	4	40	20	16	31	39	29	100	72
Teacher Education	60	99	24	57	4	61	22	17	24	34	28	40	51
Other Education	92	88	68	62	47	33	52	25	54	66	63	39	88
Aeronautical Engineering	91	57	57	48	57	24	32	23	33	21	44	49	87
Biomedical Engineering	94	63	68	50	46	31	31	24	35	23	44	69	99
Chemical Engineering	100	60	80	44	23	35	32	35	29	27	44	48	86
Civil Engineering	97	54	61	60	43	29	37	46	39	29	39	44	88
Computer Engineering	80	60	63	44	100	29	19	12	33	23	44	27	86
Electrical Engineering	84	58	63	46	73	30	27	25	32	22	43	45	88
Mechanical Engineering	94	58	72	51	48	31	38	37	30	25	43	56	84
Systems Engineering	94	65	86	57	68	33	43	34	32	32	50	56	83
Other Engineering	83	61	74	54	57	36	34	35	33	31	44	44	83
Engineering Technology	85	57	77	56	37	28	39	40	32	41	40	62	89
Foreign Lang & Linguistics	61	90	30	39	23	16	27	15	44	17	28	30	84
Family & Consumer Sciences	64	95	21	60	5	73	20	20	21	36	25	38	50
Legal Studies	69	67	44	66	15	40	38	54	50	33	42	33	74
English, Liberal Arts, Humanities	73	84	40	60	26	36	44	26	60	25	44	32	75
Library Science	78	79	43	65	40	31	46	31	49	38	48	39	80
Biochem & Molecular Biology	99	64	44	55	14	21	32	17	35	16	49	87	97
Microbiology	100	58	69	49	13	25	36	29	32	29	39	77	90
Biology	91	61	54	51	24	29	35	26	36	27	41	69	93
Statistics	97	74	69	55	75	39	55	34	37	26	51	26	84
Mathematics	92	66	67	53	78	34	42	28	37	27	47	27	82
Fitness & Leisure Studies	49	74	37	53	17	50	34	26	26	41	41	55	77
Philosophy & Religion	70	74	35	46	21	19	22	23	36	31	34	30	70
Theology	31	68	15	38	3	51	21	12	20	22	36	27	47
Atmospheric Sci & Meteorology	63	64	26	44	25	15	24	11	52	17	33	45	100

Note: Table presents the percent of each major's postings that list each skill composite. Mean and standard deviation are calculated equally weighting 70 majors. Authors' analysis of Burning Glass job postings data.

Continued: Share of Ads for Each Major Indicating Demand for Each Skill Composite

Major	Cognitive	Social	Project Mgmt	Organizational	Software	Customer Service	Computer Service	Financial	Writing	People Mgmt	Communication	Other Skills (> top 1000)	Other Skills (< top 1000)
Chemistry	100	57	65	49	15	30	36	27	33	27	42	60	87
Geological & Earth Sciences	89	53	60	58	27	30	30	37	46	35	35	55	94
Physics	100	58	60	43	67	29	24	18	34	24	41	37	83
Materials Science & Eng	94	62	72	43	25	31	31	26	30	23	47	90	87
Other Physical Sciences	90	53	56	54	27	22	22	25	38	41	35	56	89
Psychology	87	79	42	55	17	58	36	22	34	44	39	50	74
Protective Services	72	59	50	50	23	28	33	36	40	35	33	72	84
Public Administration	75	69	79	70	23	38	43	67	49	55	36	100	76
Public Policy	86	85	71	73	28	39	49	45	67	38	59	46	83
Social Work	70	74	34	54	4	78	32	21	31	38	32	54	64
Economics	100	75	68	64	45	44	60	61	39	30	52	30	79
Geography	82	62	50	61	72	35	41	20	50	20	42	31	97
Poli Sci/Gov & Intl Relations	82	80	56	68	25	35	45	40	60	37	49	47	78
Sociology	96	76	42	58	14	65	38	26	37	48	34	58	74
Other Social Sciences	86	72	50	63	30	32	37	31	51	31	38	41	91
Applied Arts	94	87	52	66	77	45	40	22	36	17	46	39	92
Other Visual/Performing Arts	76	83	37	66	61	29	32	19	59	18	42	51	95
Health & Medical Admin Services	75	69	84	58	26	67	45	53	37	51	44	47	75
Allied Health	52	56	38	38	8	67	23	18	18	30	27	82	96
Mental & Social Health Services	57	98	28	43	4	75	27	13	26	39	25	65	68
Pharm Sciences & Admin	75	74	67	50	13	55	35	35	38	38	52	51	85
Public Health	77	74	98	58	22	48	44	39	44	43	46	53	84
Rehab & Therapeutic Professions	56	67	34	46	4	76	19	27	22	67	29	54	87
Dietetics & Nutrition Services	42	67	36	58	6	60	33	26	18	31	28	54	91
Nursing	47	60	31	49	4	82	23	16	14	36	30	70	62
Other Allied Health	72	64	73	51	22	61	39	39	29	43	41	58	75
Accounting	73	61	52	62	35	33	62	92	30	28	46	28	68
Finance	82	68	62	64	40	39	63	82	32	29	50	30	71
Hospitality Admin/Mgmt	59	74	75	68	9	64	47	61	27	65	41	54	62
Human Resources Mgmt & Services	69	81	66	66	37	33	60	43	36	76	55	31	73
Marketing	79	89	67	69	33	84	52	37	49	35	56	30	79
Construction Mgmt	77	64	100	79	29	33	59	70	34	37	43	41	76
Mgmt Info Systems & Science	88	68	78	57	96	45	38	31	40	36	50	29	81
Business	78	77	77	65	40	56	51	56	36	43	53	35	75
Minimum	31	43	15	38	1	15	19	11	12	16	20	25	40
Maximum	100	99	100	87	100	84	63	92	100	76	63	100	100
Mean	79	70	56	57	33	42	38	34	38	34	42	49	81
Standard Deviation	15	12	19	10	24	17	12	17	14	12	9	18	12

Note: Table presents the percent of each major's postings that list each skill composite. Mean and standard deviation are calculated equally weighting 70 majors. Authors' analysis of BGT job postings data.

Table B.8: Correlation between Different Measures of Major Skill Specificity

	Outcome							
	A: Similarity based on 9000 skills				B: LQ measure			
	Rank		Measure		Rank		Measure	
Outcome Weights	NO	YES	NO	YES	NO	YES	NO	YES
LQ measure (only top 1000 skills)	.37	.53	.41	.57				
Similarity (Full)					.37	.53	.41	.57
Similarity (top 1000)	.90	.96	.90	.99	.36	.58	.39	.58
Similarity (1001+)	.32	.47	.30	.56	.17	.37	.20	.37
% of recent grads in top 5 occs	.05	.32	.08	.34	.00	.47	.02	.47

Note: "Full similarity" is the cosine similarity (or rank) of a major using all 9000 skills. Top 1000 is the cosine similarity using only the 1000 most frequent skills. 1001+ is cosine similarity using skills ranked 1001-9000 in terms of overall frequency. LQ is location quotient across 11 skill composites (calculated as sum(abs(LQ-1) across the composites) and expressed in either rank or actual measure. Percent of recent graduates in top 5 occupations measures the fraction of a major's graduates aged 23-27 that are found in the 5 most frequent occupations for the major in the ACS. Panel A regresses a major's rank (measure) for the full similarity on the rank (measure) of the variable in the first column. Panel B does the same but with outcomes based on sum(abs(LQ-1)). Each regression has 70 observations (1 for each major) except for % in top 5 occupations which has 66 observations because 4 majors are missing from the ACS. Each cell is the adjusted R-squared from the regression. In weighted regressions, majors are weighted by the number of job postings.

Table B.9: Comparison of Major Rankings by Measure of Specificity

Most specific majors (top 10)		
LQ-based rank	Cosine-based rank	Gini-based rank
Culinary Arts	Family & Consumer Sciences	Primary/General Education
Nursing	Special Education & Teaching	Secondary Education
Special Education & Teaching	Mental & Social Health Services	Nursing
Allied Health	Teacher Education	Medical Tech
Rehab & Therapeutic Professions	Atmospheric Science & Meteorology	Computer Programming
Mental & Social Health Services	Culinary Arts	Other Med/Health Services
Theology	Microbiology	Finance
Foreign Language & Linguistics	Rehab & Therapeutic Professions	Precision Production/Industrial Arts
Biochem & Molecular Biology	Biochem & Molecular Biology	Commercial Art and Design
Atmospheric Science & Meteorology	Allied Health	Marketing
Most general majors (top 10)		
LQ-based rank	Cosine-based rank	Gini-based rank
Other Engineering	Business	Music/Speech/Drama
Architecture	Other Engineering	Other Social Sciences
Civil Engineering	Marketing	Philosophy/Religion
Business	Other Allied Health	Environmental Studies
Economics	Library Science	Psychology
Mathematics	Health & Medical Admin Services	Accounting
Urban Planning	Pharmacy Sciences & Administration	Area Studies
Systems Engineering	Legal Studies	Social Work/Human Resources
Mechanical Engineering	Mathematics	Mathematics
Management Information Systems & Science	Political Science, Government, International Relations	Engineering Tech

Note: This table mirrors the layout of Table 3 in Leighton and Speer (2020), comparing the top and bottom 10 majors in terms of specificity based on different measures: thus, majors in the "Most specific" panel are listed from most specific to least specific; majors in the "Most general" panel are listed from least specific (i.e., most general) to more specific. Our two ranking measures appear in italics. Rankings in the Gini-based column come from Table 3 in Leighton and Speer (2020).

Table B.10: Major Specific Skill Similarity Measures

Major Name	% of all postings	% of ads x major	cosine similarity	LQ norm measure 1	LQ norm measure 2
Accounting	13.87	8.22	0.73	3.29	1.94
Aeronautical Engineering	0.44	0.26	0.73	2.69	0.86
Agriculture	0.82	0.48	0.78	2.05	0.96
Allied Health	0.10	0.06	0.51	5.39	3.50
Applied Arts	1.01	0.60	0.59	2.43	0.94
Architecture	0.34	0.20	0.70	1.41	0.29
Atmospheric Sci & Meteorology	0.03	0.02	0.45	4.57	2.37
Biochem & Molecular Biology	0.18	0.11	0.51	4.58	3.28
Biology	1.40	0.83	0.72	3.01	1.36
Biomedical Engineering	0.19	0.11	0.62	2.64	1.17
Business	29.54	17.51	0.96	1.76	0.38
Chemical Engineering	0.61	0.36	0.56	2.64	0.75
Chemistry	1.77	1.05	0.57	2.97	1.25
Civil Engineering	0.95	0.57	0.57	1.63	0.32
Communication & Media Studies	2.57	1.52	0.82	3.04	1.51
Computer & Info Science	26.15	15.50	0.79	2.70	1.38
Computer Engineering	2.48	1.47	0.55	3.70	2.20
Construction Mgmt	0.91	0.54	0.63	2.91	1.24
Culinary Arts	0.19	0.11	0.46	6.46	5.64
Dietetics & Nutrition Services	0.29	0.17	0.59	3.77	1.95
Economics	3.29	1.95	0.73	1.91	0.54
Electrical Engineering	5.73	3.40	0.82	2.62	0.84
Engineering Technology	0.88	0.52	0.80	2.16	0.74
English, Liberal Arts, Humanities	0.14	0.08	0.84	2.96	1.21
Family & Consumer Sciences	0.38	0.22	0.39	4.35	2.54
Finance	11.15	6.61	0.83	2.38	1.24
Fitness & Leisure Studies	0.37	0.22	0.81	3.25	1.30
Foreign Lang & Linguistics	0.11	0.07	0.63	4.60	2.19
Geography	0.17	0.10	0.68	2.64	0.96
Geological & Earth Sciences	0.48	0.28	0.59	2.44	0.79
Health & Medical Admin Services	0.95	0.56	0.86	2.41	0.92
Hospitality Admin/Mgmt	0.26	0.15	0.81	4.02	2.29
Human Resources Mgmt & Services	2.08	1.23	0.82	2.92	2.09
Journalism	1.15	0.68	0.60	4.15	4.11
Legal Studies	0.73	0.43	0.85	2.39	0.95

Note: Table displays three different measures of a major's skill specificity using the skills listed on job postings in the Burning Glass data. For each major, cosine similarity is constructed using a vector of the share of ads listing each of the 9,000 most common skills for each major and for all job postings in the analysis sample. The LQ norm measure 1 is calculated as the sum across all 11 skill composites of the absolute value of the deviations of the LQs from 1, and LQ norm measure 2 is calculated as the sum of the squared deviations.

Continued: Major Specific Skill Similarity Measures

Major Name	% of all postings	% of ads x major	cosine similarity	LQ norm measure 1	LQ norm measure 2
Library Science	0.11	0.07	0.87	2.07	0.56
Marketing	5.57	3.30	0.88	2.72	1.20
Materials Science & Eng	0.17	0.10	0.58	3.94	2.68
Mathematics	2.20	1.31	0.85	1.98	0.63
Mechanical Engineering	4.29	2.54	0.74	2.02	0.52
Mental & Social Health Services	0.07	0.04	0.41	5.28	3.10
Mgmt Info Systems & Science	4.49	2.66	0.75	2.04	1.05
Microbiology	0.44	0.26	0.50	3.49	2.02
Natural Resources	0.35	0.21	0.71	2.55	1.08
Nursing	8.42	5.00	0.62	5.53	3.63
Other Allied Health	2.40	1.42	0.88	2.43	0.87
Other Education	0.28	0.17	0.72	3.05	1.63
Other Engineering	16.46	9.76	0.92	1.39	0.21
Other Physical Sciences	0.03	0.02	0.61	3.21	1.23
Other Social Sciences	0.11	0.06	0.76	2.15	0.63
Other Visual/Performing Arts	0.10	0.06	0.62	3.65	1.56
Pharm Sciences & Admin	0.23	0.14	0.86	2.16	0.82
Philosophy & Religion	0.02	0.01	0.78	3.30	1.45
Physics	0.89	0.53	0.57	2.81	1.01
Poli Sci/Gov & Intl Relations	0.33	0.20	0.85	2.43	0.97
PR & Advertising	1.01	0.60	0.80	3.31	1.69
Protective Services	0.11	0.07	0.70	2.95	1.40
Psychology	1.41	0.84	0.66	2.84	1.11
Public Administration	0.77	0.46	0.63	4.41	3.90
Public Health	0.92	0.54	0.74	2.28	0.88
Public Policy	0.16	0.09	0.84	2.75	1.28
Rehab & Therapeutic Professions	0.31	0.19	0.51	5.31	3.41
Social Work	1.56	0.93	0.62	3.81	2.12
Sociology	0.39	0.23	0.61	3.28	1.47
Special Educ & Teaching	0.22	0.13	0.41	5.45	4.82
Statistics	1.68	1.00	0.78	2.14	0.63
Systems Engineering	0.68	0.40	0.82	1.99	0.60
Teacher Education	0.53	0.31	0.44	4.05	2.31
Theology	0.07	0.04	0.72	5.09	3.14
Urban Planning	0.24	0.14	0.72	1.98	0.61

Note: Table displays three different measures of a major's skill specificity using the skills listed on job postings in the Burning Glass data. For each major, cosine similarity is constructed using a vector of the share of ads listing each of the 9,000 most common skills for each major and for all job postings in the analysis sample. The LQ norm measure 1 is calculated as the sum across all 11 skill composites of the absolute value of the deviations of the LQs from 1, and LQ norm measure 2 is calculated as the sum of the squared deviations.

Table B.11: Replication and Extension of Deming and Kahn (2018)

		Replication: Occupation-MSA Cell				Our sample: Major-MSA Cells			
		All education levels		Education = 16		Education = 16		Education = 16	
	DK estimates	Keyword	Hand-coded	Keyword	Hand-coded	Keyword	Hand-coded	Keyword	Hand-coded
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share cognitive	0.0792***	0.0484	0.0601*	0.0359	0.0593*	-0.0021	0.125*	0.0076	-0.0049
		(0.0357)	(0.0341)	(0.0303)	(0.0345)	(0.0519)	(0.0685)	(0.0193)	(0.0246)
Share social	0.0517***	0.0508*	-0.0129	0.0566*	0.0174	0.0642	0.0509	-0.0123	0.0090
		(0.0264)	(0.0385)	(0.0328)	(0.0149)	(0.0422)	(0.0438)	(0.0169)	(0.0199)
Observations		54,216	54,216	43,848	43,848	22,151	22,151	22,151	22,151
Major & MSA FE	X	X	X	X	X	X	X	X	X
Addt'l Controls	X	X	X	X	X	–	–	–	–
Outcome		log(mean hourly wage) from OES				log(mean hourly wage) from ACS			
Weights		Job postings from BG				Job postings from BG		Person wt from ACS	

Note: Outcome is the average of log of mean hourly earnings (2019 dollars) among college graduates as measured in the American Community Survey (ACS) or from the Occupation Employment and Wage Statistics (OES). Each observation is a major-msa or occupation-msa cell. DK estimates are from Table 3 column 5 of Deming and Kahn (2018). Addt'l controls includes 6-digit occ FE, % of postings in 2 digit industry, education and experience. All models also include the share of ads in each cell that require customer service, financial, organizational, people management, project management, writing, basic computer, and software skills.

Table B.12: Raw Cell-level Correlations between Social and Cognitive Skill Content and Wages, Robustness

	MSA x major	MSA x occ	MSA x major	MSA x occ	MSA x major	MSA x occ
	(1)	(2)	(3)	(4)	(5)	(6)
Share cognitive						
Keyword	0.855***	0.569***	0.399***	0.290***	0.746***	0.430***
	(0.0101)	(0.0089)	(0.0106)	(0.0067)	(0.0114)	(0.0090)
Hand-coded	0.498***	0.359***	0.417***	0.187***	0.745***	0.347***
	(0.0122)	(0.0092)	(0.0112)	(0.0068)	(0.0122)	(0.0094)
Share social						
Keyword	0.205***	0.688***	0.141***	0.257***	0.394***	0.875***
	(0.0130)	(0.0104)	(0.0120)	(0.0072)	(0.0146)	(0.0113)
Hand-coded	-0.601***	0.0222**	-0.282***	-0.0459***	-0.464***	0.369***
	(0.0126)	(0.0112)	(0.0121)	(0.0073)	(0.0137)	(0.0115)
Weights	ACS perwt	soc emp	none	none	postings	postings
Observations	22,151	43,852	22,151	43,852	22,151	43,852

Note: Each cell is a separate regression of cell-level log mean wages (major-MSA or occupation-MSA) on the share of ads requiring each skill. Outcome is the average of log of mean hourly earnings (2019 dollars) among college graduates as measured in the American Community Survey (ACS) or from the Occupation Employment and Wage Statistics (OES). Share cognitive (social) is the percent of job postings in major-MSA or occupation-MSA cell that demand the skill demand as measured in the Burning Glass data. Weights in Column (2) (soc emp) is total employment count from the OES.

APPENDIX C

Appendix to Chapter 3

C.1 Additional Figures & Tables

Table C.1: Descriptive Statistics

	NSCG	NSCG (weighted)	ACS (weighted)
Demographic Characteristics:			
female	40%	45%	47%
black	8%	7%	8%
hispanic	10%	7%	5%
potential experience	18	19	20.6
Bachelor's Degree Graduation Year:			
mean	1995	1994	1994
1965-1974	7%	6%	7%
1975-1984	18%	19%	20%
1985-1994	22%	24%	25%
1995-2004	24%	26%	26%
2005+	30%	25%	23%
College Major Subject Field:			
Business & Economics	10%	21%	22%
Computer Science & Engineering	32%	15%	14%
Communications & Marketing	3%	8%	7%
Education	3%	8%	10%
Health	5%	5%	6%
Humanities	6%	12%	12%
Other	3%	3%	5%
Pure Sciences	22%	13%	11%
Social Sciences	17%	15%	13%
Graduate Degrees:			
any degree	48%	36%	37%
Master's	36%	27%	27%
Professional	5%	7%	4%
Doctoral	7%	3%	6%

Note: Table presents descriptive statistics for the 2003-2019 National Survey of College Graduates (NSCG) and the 2009-2019 American Community Survey (ACS). Estimates in Column 2 are weighted using the adjusted NSCG survey weights and estimates in Column 3 are weighted using ACS person weights.

Table C.2: Three Most Common Occupations by Primary Work Activity

Work Activity	Broad Occupation (% in Occupation Reporting Task)		
accounting	Business Related (51.5%)	Managers (21.8%)	Clerical (14.9%)
basic research	Biological scientists (21.2%)	Business Related (8.4%)	Postsecondary Teachers (8.1%)
applied research	Biological scientists (13.4%)	Engineers (11.1%)	Managers (10.5%)
development	Engineers (25%)	Managers (12.6%)	Computer Scientist (10.9%)
design	Engineers (31.5%)	Computer Scientist (22.2%)	Managers (12.1%)
computer applications	Computer Scientist (53.9%)	Other Computer (16.1%)	Managers (8.3%)
employee relations	Business Related (40.4%)	Managers (24.7%)	Clerical (8.6%)
project & people mgmt	Managers (50.9%)	Business Related (6.3%)	Marketing (5.9%)
production	Blue collar (29.7%)	Managers (10.6%)	Technician (8.5%)
professional services	Other Health (24.6%)	Law Related (18.4%)	Doctors (13.4%)
sales	Marketing (42%)	Managers (16%)	Business Related (15.3%)
quality mgmt	Managers (26.6%)	Blue collar (9.5%)	Engineers (8.9%)
teaching	Primary and Secondary Teachers (80.9%)	Postsecondary Teachers (7.7%)	Other Social Service (4.3%)

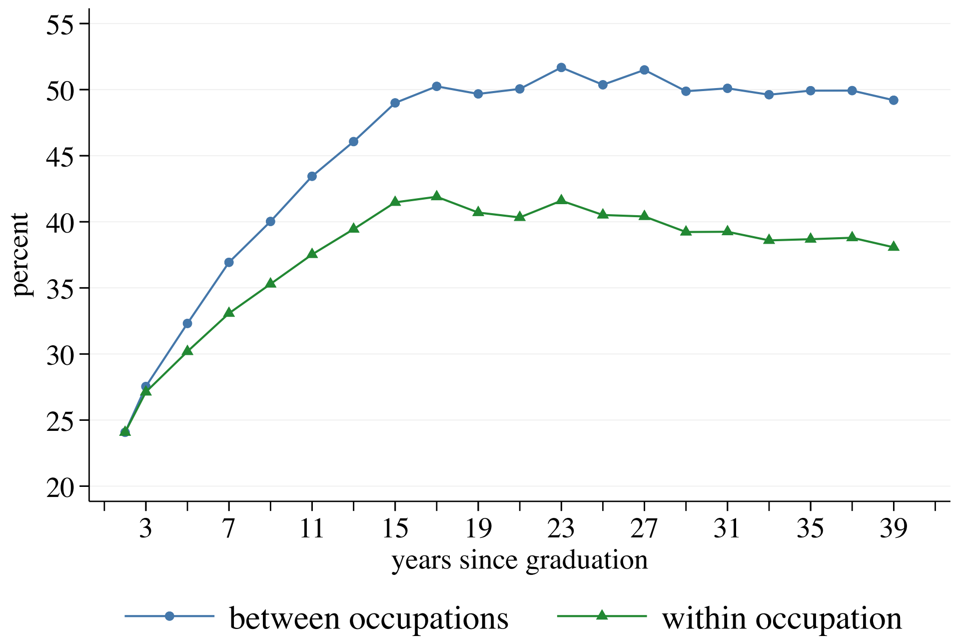
Note: For each primary work activity the table presents the three occupations with the large percent of individuals performing the primary work activity. The percent of individuals working in each occupation that report performing the primary work activity is provided in parenthesis. Primary work activity is the one in which the respondent spend the most hours on during a typical week performing. Observations are weighted using NSCG survey weights. N= 250,265.

Table C.3: Three Most Common Primary Work Activities by Occupation

Occupation	Work Activity (% in Occupation Reporting Task)		
Biological scientists	applied research (31.4%)	basic research (20.3%)	project & people mgmt (12.4%)
Blue collar	production (24.8%)	project & people mgmt (23.2%)	sales (8.3%)
Business Related	accounting (40.7%)	sales (17.7%)	professional services (12.8%)
Clerical	accounting (26.1%)	sales (24.3%)	project & people mgmt (12.8%)
Computer Scientist	computer applications (46.1%)	design (15.7%)	project & people mgmt (13.1%)
Doctors	professional services (93.2%)	applied research (1.5%)	project & people mgmt (1.4%)
Engineers	design (23.3%)	project & people mgmt (22.6%)	development (14.1%)
Farmers, Foresters & Fisherman	production (34.3%)	project & people mgmt (23.8%)	accounting (7.7%)
Law Related	professional services (87.4%)	project & people mgmt (4.1%)	accounting (3.2%)
Managers	project & people mgmt (52.2%)	sales (11.9%)	accounting (11.1%)
Marketing	sales (71.2%)	project & people mgmt (13.9%)	accounting (3.3%)
Math Scientist	applied research (37.8%)	computer applications (14.2%)	project & people mgmt (8.6%)
Other Computer	computer applications (34.8%)	project & people mgmt (14.2%)	design (13.5%)
Other Health	professional services (68.1%)	project & people mgmt (8.2%)	teaching (3.2%)
Other Service	project & people mgmt (26.8%)	sales (25.7%)	professional services (11.2%)
Other Social Service	professional services (45.5%)	project & people mgmt (16.7%)	teaching (10.2%)
Physical Scientist	applied research (29.8%)	project & people mgmt (13.3%)	basic research (11.9%)
Postsecondary Teachers	teaching (64.9%)	professional services (9.1%)	project & people mgmt (7.9%)
Primary and Secondary Teachers	teaching (91.7%)	project & people mgmt (3.2%)	professional services (1%)
Social Scientist	professional services (36.2%)	applied research (20.4%)	project & people mgmt (10.8%)
Technician	production (23.3%)	project & people mgmt (15.7%)	applied research (11.4%)
Writers and Artists	project & people mgmt (15.8%)	professional services (15.3%)	sales (14.7%)

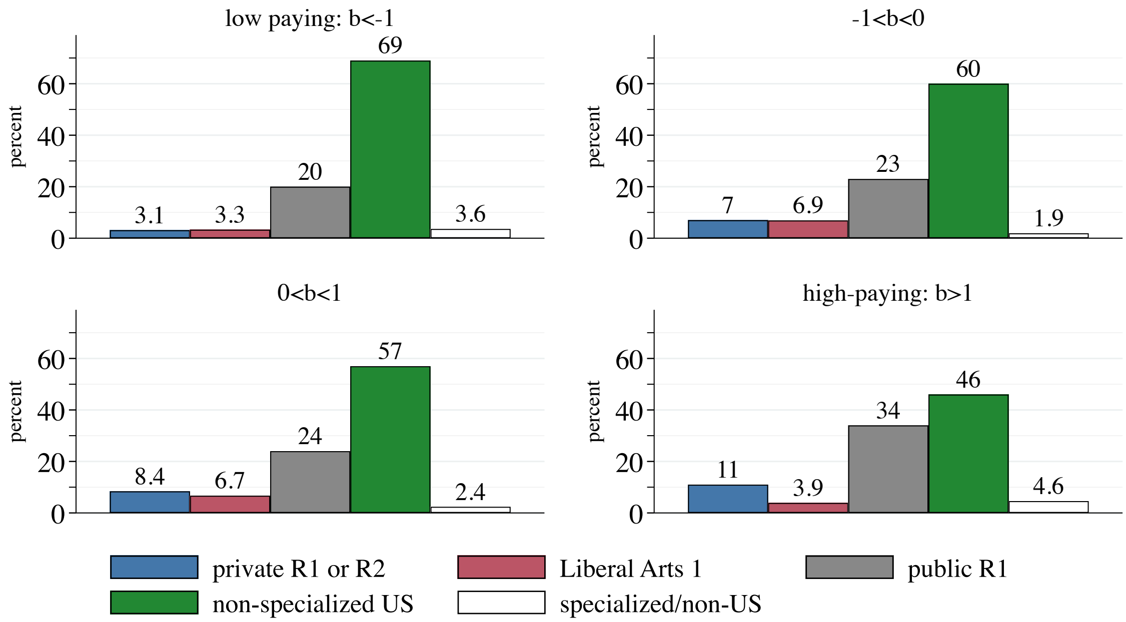
Note: For each occupation the table presents the three most common primary work activities. The percent of individuals working in each occupation that report performing the primary work activity is provided in parenthesis. Primary work activity is the one in which they spend the most hours on during a typical week performing. Observations are weighted using NSCG survey weights. There are 9,803 observations that do not report any primary work activity. N= 240,462.

Figure C.1: Percent of Workers Supervising Others by Years Since Graduation



Note: Figure plots the predicted percent of individuals performing supervisory work in each year since graduation bin. The coefficients on years since graduation bins are estimated using a regression with a dummy for performing supervisory work as the outcome and controls including the survey year fixed effects, indicators for female, black, Hispanic, an indicator for parental college education, and either with occupation fixed effects (*within occupations*) or without (*between occupations*). Observations weighted using survey-weights. Data source is the 2003-2019 NSCG. N=250,265 in each regression.

Figure C.2: Distribution of Institution Type by Standardized Major Fixed Effect



Note: Figure plots the percent of observations in each category of the standardized major fixed effect with each undergraduate institution type. Undergraduate institution is defined similarly to Hersch (2013) using the 1994 Carnegie Classification. See text for details. The 61 majors are divided into four categories based on the standardized major fixed effect $\hat{b}_{m,std}$ (b in the figure): $\hat{b}_{m,std} < -1$, $-1 < \hat{b}_{m,std} < 0$, $0 < \hat{b}_{m,std} < 1$ and $\hat{b}_{m,std} > 1$. The standardized major fixed effect is based on the major fixed effects from a regression of log annual earnings on 61 college major dummies and survey year fixed effects, a quartic in years since first bachelor's graduation, indicators for female, black, Hispanic, an indicator for parental college education, high school region and 10-year graduation cohort fixed effects. The major fixed effects are then standardized fixed effects are calculated using the mean and standard deviation of the 61 major fixed effects (with survey-weighted employment). Observations are weighted using NSCG survey weights. Data source is the 2003-2019 NSCG. N=250,265.

Table C.4: Major Fixed Effects & Standardized Major Fixed Effects

	base model		full model	
	major FE	standardized	major FE	standardized
Accounting	.4673***	1.045	.2081***	1.328
Aeronautical Engineering	.5211***	1.364	.2439***	1.767
Agriculture	.1170***	-1.032	.0497***	-0.610
Allied Health	.2490***	-0.249	.0699***	-0.362
Applied Arts	.0103	-1.665	-.0374**	-1.675
Architecture	.2510***	-0.237	.1054***	0.072
Atmospheric Sciences and Meteorology	.2134***	-0.460	.0657*	-0.414
Biochemistry, Biophysics and Molecular Biology	.4936***	1.201	.1658***	0.812
Biology	.3825***	0.542	.0920***	-0.092
Biomedical Engineering	.5346***	1.444	.1485***	0.600
Business, general	.2639***	-0.161	.0872***	-0.150
Chemical engineering	.5816***	1.723	.2603***	1.968
Chemistry	.4567***	0.982	.1628***	0.775
Civil Engineering	.4489***	0.936	.2085***	1.334
Communications	.1759***	-0.682	.0410***	-0.715
Computer and Information Sciences, General	.4520***	0.955	.1722***	0.890
Computer Engineering	.6035***	1.853	.2548***	1.900
Economics	.4699***	1.060	.1920***	1.132
Electrical, Electronics, Communications Engineering	.5373***	1.460	.2286***	1.580
Engineering technology	.2939***	0.017	.1272***	0.338
English, Liberal Arts, Humanities	.1981***	-0.551	.0511***	-0.592
Family and Consumer Sciences	.0488*	-1.436	-.0279	-1.559
Finance	.4482***	0.932	.2001***	1.231
Fitness, Recreation and Leisure Studies	.0395*	-1.491	-.0468**	-1.790
Foreign Language & Linguistics	.2019***	-0.528	.0516***	-0.586
Geography	.1685***	-0.726	.0383*	-0.749
Geology and Earth Science	.2234***	-0.401	.0707***	-0.352
Health and Medical Administrative Services	.2719***	-0.113	.0867***	-0.157
Industrial And Manufacturing Engineering	.5666***	1.634	.2779***	2.183
Journalism	.2045***	-0.513	.0817***	-0.218
Legal Studies	.2759***	-0.090	.0373	-0.761
Library Science	.0756	-1.277	.0717	-0.340

Note: Table percent the major fixed effects $\hat{\beta}_m$ and the standardized major fixed effects for the baseline and the full model. Baseline major fixed effects come from the baseline model Equation (3.3.1) and the full model fixed effects come from Equation (3.3.2). The standardized major fixed effect for the base model are $\left[\hat{\beta}_m^{base} - \overline{\hat{\beta}_m^{base}}\right] / \left[\sigma(\hat{\beta}_m^{base})\right]$ and for the full model are $\left[\hat{\beta}_m^{full} - \overline{\hat{\beta}_m^{full}}\right] \left[\sigma(\hat{\beta}_m^{full})\right]$. The major fixed effects are then standardized fixed effects are calculated using the mean and standard deviation of the 61 major fixed effects (with survey-weighted employment). Observations are weighted using NSCG survey weights. Data source is the 2003-2019 NSCG. N=250,265.

Continued: Major Fixed Effects & Standardized Major Fixed Effects

	base model		full model	
	major FE	standardized	major FE	standardized
Management Information Systems and Science	.6915***	2.375	.2851***	2.270
Marketing	.2974***	0.038	.1352***	0.436
Materials Science and Engineering	.4576***	0.988	.1663***	0.818
Mathematics	.3502***	0.351	.1277***	0.346
Mechanical engineering	.5039***	1.262	.2250***	1.536
Microbiology	.4351***	0.855	.1777***	0.957
Natural Resources	.1120***	-1.061	.0180	-0.997
Nursing	.4066***	0.685	.1331***	0.411
Nutritional sciences	.1897***	-0.601	-.0447	-1.764
Other Education	–	-1.725	–	-1.217
Other Engineering	.5016***	1.249	.2356***	1.666
Other Physical Sciences	.2160***	-0.445	.0519	-0.583
Other Social Sciences	.1806***	-0.654	.0204*	-0.968
Other Visual/Performing Arts	-.0088	-1.778	-.0493***	-1.820
Pharmacy	.6224***	1.965	.2926***	2.362
Philosophy, Religion & Theology	-.0445*	-1.989	-.0801***	-2.197
Physics	.4247***	0.792	.1635***	0.783
Political Science, Government, Int'l Relations	.3982***	0.636	.1120***	0.153
Protective Services	.1145***	-1.047	.0423**	-0.700
Psychology	.1831***	-0.640	.0362***	-0.774
Public Administration	.2902***	-0.005	.0977**	-0.022
Public Health	.1888***	-0.606	.0714***	-0.343
Public Policy	.4110***	0.711	.1727***	0.896
Rehabilitation and Therapeutic Professions	.2110***	-0.474	.0234*	-0.931
Social Work	.0255	-1.574	.0015	-1.199
Sociology	.1286***	-0.963	.0225**	-0.942
Special Education and Teaching	.0565**	-1.391	.0455*	-0.661
Statistics	.4133***	0.725	.1539***	0.665
Teacher Education	.0354**	-1.516	.0131	-1.057

Note: Table percent the major fixed effects $\hat{\beta}_m$ and the standardized major fixed effects for the baseline and the full model. Baseline major fixed effects come from the baseline model Equation (3.3.1) and the full model fixed effects come from Equation (3.3.2). The standardized major fixed effect for the base model are $\left[\hat{\beta}_m^{base} - \overline{\hat{\beta}_m^{base}}\right] / \left[\sigma(\hat{\beta}_m^{base})\right]$ and for the full model are $\left[\hat{\beta}_m^{full} - \overline{\hat{\beta}_m^{full}}\right]\left[\sigma(\hat{\beta}_m^{full})\right]$. The major fixed effects are then standardized fixed effects are calculated using the mean and standard deviation of the 61 major fixed effects (with survey-weighted employment). Observations are weighted using NSCG survey weights. Data source is the 2003-2019 NSCG. N=250,265.

Table C.5: Fraction of Variation in Work Activities Explained by Major and Occupation

	Variation in skill-share explained by...			
	(1) Major	(2) Occupation	(3) Major & Occ	Unexplained
accounting	0.15	0.32	0.34	0.66
basic research	0.02	0.08	0.08	0.92
applied research	0.02	0.11	0.11	0.89
development	0.02	0.05	0.05	0.95
design	0.05	0.12	0.12	0.88
computer applications	0.15	0.30	0.31	0.69
employee relations	0.01	0.18	0.18	0.82
project & people mgmt	0.02	0.19	0.20	0.80
production	0.02	0.14	0.15	0.85
professional services	0.15	0.52	0.52	0.48
sales	0.05	0.34	0.35	0.65
quality mgmt	0.01	0.02	0.02	0.98
teaching	0.19	0.77	0.77	0.23

Note: An indicator for whether individual i performs task k is regressed on major fixed effects (Column 1), occupation fixed effects (Column 2) and major and occupation fixed effects (Column 3). Reported is the R^2 for the regression. Observations are weighted NSCG survey weights. The amount of variation unexplained by major and occupation is equal to 1 minus the r-squared in Column 3. Data source is the 2003-2019 NSCG. N= 250,265.

C.2 Data

Occupation I work with two different aggregates of occupation codes: 83 detailed occupations and 20 broad occupations. To create the codes I first harmonize occupation codes across the American Community Survey (ACS) waves, and then across the ACS and National Survey of College Graduates (NSCG) surveys.

The ACS codes found in IPUMs vary a bit between the 2009, 2010-2011, 2012-2017, 2018-2019 survey waves. I first crosswalk all ACS codes to the 2010 ACS codes. The largest adjustments were between the 2009 ACS codes which are based in the 2000 Census and SOC occupation coding scheme and the 2010 ACS codes which are based in the 2010 Census and SOC codes. I used the general rules to aggregate codes. If a single 2009 ACS codes mapped to several 2010+ codes, I aggregated the 2010+ codes into a single occupation and assigned it the single 2009 code. If several 2009 codes mapped to a single 2010+ code, I aggregated the 2009 codes. The 2018 and 2019 splits several of the occupation codes found in 2010-2017 ACS and I collapse the 2008 to 2019 codes back into their 2010 codes.

I next aggregated the ACS occupation codes into the NSCG occupation codes based on the names of the occupations. For some large classes of occupations that entail more routine-manual work (e.g. Precision/production occupations and Transportation and Moving Materials occupations) I use David Dorn's crosswalk of the occ2010 to occ1990dd codes. The result is a system of 83 occupations that are harmonized across the ACS and NSCG. Finally, I aggregate the occupation codes into roughly 20 broad occupation codes following Altonji and Zhong (2021). As the NSCG occupation codes vary in their detail across broad groups of occupations, I primarily focus on outcomes based on the broad occupation codes.[1] The broad and narrow occupation codes include:

- Biological scientists: Agricultural and food scientists, Biological scientists, Foresters and conservation scientists, Medical and Life Scientists
- Blue Collar Occupations: Construction and extraction occupations; Installation, maintenance, and repair occupations; Precision/production occupations; Protective services (e.g., fire fighters, police, guards, wardens, park rangers); Transportation and Moving Materials occupations
- Business Related Occupations: Accountants and Auditors and other financial specialists; Actuaries; Business and Financial Operations Occupations (Insurance securities real estate and business services); Personnel training and labor relations specialists
- Clerical Occupations: Accounting clerks and bookkeepers; Other Office and Administrative Support Occupations; Secretaries, receptionists, typists
- Computer Scientists: Computer Network Architects; Computer programmers (business, scientific, process control); Computer system analysts; OTHER computer information science occupations; Operations and systems researchers and analysts; Software developers
- Doctors: Diagnosing/treating practitioners (Physicians, Dentists, Veterinarians, Optometrists, Podiatrists)
- Engineers: Aerospace Engineers; Architects; Biomedical and Agricultural Engineers; Chemical Engineers; Civil Engineers; Computer Hardware Engineers; Electrical and

[1] For example Computer Occupations are very disaggregated and include Computer programmers (business, scientific, process control, Computer system analysts, Computer support specialists, Database administrators, Information security analysts, Network and computer systems administrators, other computer information science occupations, Software developers, Web developers and Computer Network Architects. Health Occupations are fairly aggregated with occupations including "Diagnosing/treating practitioners: Physicians, Dentists, Veterinarians, Optometrists, Podiatrists" and "Registered nurses, pharmacists, dieticians, therapists, physician assistants, nurse practitioners". For Business occupations there is a fairly general occupation titled "Business and Financial Operations Occupations (Insurance securities real estate and business services)" but also more specific occupations like "Accountants and Auditors and other financial specialists" and "Chief Executives and Legislators and Top Level Managers".

Electronics Engineers; Environmental Engineers; Marine Engineers and Naval Architects; Materials Engineers; Mechanical Engineers; Other Engineers; Petroleum, mining and geological engineers; Sales engineers

- Farmers, Foresters and Fisherman: Farmers, Foresters and Fisherman
- Law Related Occupations: Lawyers and judges
- Managers: Chief Executives and Legislators and Top Level Managers; Education administrators (e.g. registrar dean principal); Medical and Health Services Managers; Natural Science Managers; Other Managers
- Marketing: Other Sales, Marketing and Related Occupations; Commodities sales (e.g. machinery, equipment, supplies); Retail Sales (e.g. furnishings, clothing, motor vehicles, cosmetics)
- Math Scientists: Mathematicians and statisticians
- Other Computer occupations: Computer support specialists; Database administrators; Information security analysts; Network and computer systems administrators; Web developers
- Other Health Occupations: Health technologists and technicians (e.g dental hygienists, licensed practical nurses, medical/laboratory technicians); Other Health Occupations; Registered nurses, pharmacists, dietician, therapists, physician assistants, nurse practitioners
- Other Service Occupations: Food preparation and service (e.g., cooks, waitresses, bartenders); Other service occupations, except health
- Other Social Service Occupations: Clergy and religious workers; Counselors; Librarians, Archivists, Curators; Social Workers
- Physical Scientist: Atmospheric and space scientists; Chemists and Materials Scientists; Environmental Scientists and Geoscientists; Other Physical Scientists; Physicists and Astronomists
- Postsecondary Teachers: Postsecondary Teachers
- Primary and Secondary Teachers: Education Workers, Other; Elementary and Middle School Teachers; Preschool and Kindergarten Teachers; Secondary School Teachers; Special Education Teachers
- Social Scientist: Economists; Psychologists; Social Scientists
- Technician: Drafters; Engineering Technologists/Technicians/Surveyors; Life, Physical, and Social Science Technicians; Surveyors, Cartographers, and Photogrammetrists
- Writers and Artists: Arts, Design, Entertainment, Sports, and Media Occupations

Graduate Education Graduate degree types are pre-defined in the ACS and NSCG as follows: Master's (MA) degrees include Master of Science (MS), Master of Arts (MA), and Master of Business Administration (MBD). Doctorate includes Doctor of Philosophy (PhD), Doctor of Science (DSc), and Education Doctorate Degree (EdD). Professional degrees include Juris Doctor (JD), Bachelor of Laws (LLB), Doctor of Medicine (MD), Doctor of Dental Surgery (DDS), Doctor of Veterinary Medicine (DVM).

Adjusting Survey Weights I adjust the survey weights to account for varying sample sizes across surveys while maintaining the relative weights within each survey following Altonji and Zhong (2021). For each survey s and individual i the initial survey weight $weight_{is}$ is adjusted as:

$$\text{adjusted weight}_{is} = \frac{\text{weight}_{is}}{\sum_{i=1}^{N_s} \text{weight}_{is}} N_s$$

The adjusted weights are trimmed using 1/10 and 10 times of the median of the adjusted weights.

Job Tasks (Work Activities) The definition of work activities in the NSCG are:

- *accounting, finance, contracts*
- *basic research*: study directed toward gaining scientific knowledge primarily for its own sake
- *applied research*: study directed toward gaining scientific knowledge to meet a recognized need
- *development*: using knowledge gained from research for the production of materials and devices
- *design*: design of equipment, processes, structures, models
- *computer applications*: computer programming, systems or applications development
- *employee relations*: including recruiting, personnel development, training
- *managing*: managing or supervising people or projects
- *production & operations*: production, operations, maintenance (e.g., chip production, operating lab equipment)
- *professional services*: including health care, counseling, financial services, legal services
- *sales & service*: sales, purchasing, marketing, customer service, public relations
- *quality or productivity management*
- *teaching*

C.3 Algebraic Appendix

C.3.1 Proof that $\hat{\alpha}_m^{base} = \sigma(\hat{\beta}_m^{base})$

I estimate two baseline specifications:

$$log(earn_{imt}) = b_0 + \sum_m \beta_m{}^{base}(D_m) + \beta_x X_{imt} + \epsilon \tag{C.1}$$

$$log(earn_{imt}) = \alpha_0 + \alpha_1 \hat{b}_{m,std} + \alpha_x X_{imt} + \epsilon \tag{C.2}$$

where I replace the 61 major dummies D_m in Equation (C.1) with the single standardized major fixed effect $\hat{b}_{m,std}$ which is equal to:

$$\hat{b}_{m,std} = \frac{\hat{\beta}_m - \overline{\hat{\beta}_m}}{\sigma(\hat{\beta}_m)}, \qquad \bar{\hat{\beta}}_m = \frac{\sum_m \hat{\beta}_m}{m} \qquad \sigma(\hat{\beta}_m) = \sqrt{\frac{\sum_m(\hat{\beta}_m - \bar{\hat{\beta}}_m)^2}{m-1}} \tag{C.3}$$

In what follows I show that the estimated coefficient $\hat{\alpha}_1$ on $\hat{b}_{m,std}$ estimated in Equation (C.2) is mechanically equivalent to the standard deviation of the baseline major fixed effects estimated in Equation (C.1): $\hat{\alpha_1} = \sigma(\hat{\beta}_m^{base})$.

For simplicity, let there be m majors each with one observation so that $n_m = 1$ and $m = n$. Denote log earnings with y. In the regression used to estimate the major fixed effects $\hat{\beta}_m = \bar{y}_m - \bar{y}_o$ as it is the difference between average earnings of major m over the average earnings of the omitted major o. The expression for the coefficient $\hat{a}_1$ on $\hat{b}_{m,std}$ from model $y = a_0 + a_1 \hat{b}_{m,std} + a_x X_{imt} + \epsilon$ is:

$$\hat{a}_m = \frac{cov(\hat{b}_{m,std}, y)}{\text{Var}(\hat{b}_{m,std})} = \frac{\sum_m(\hat{b}_{m,std} - \overline{\hat{b}_{m,std}})(y_m - \bar{y})}{(m-1)\text{Var}(\hat{b}_{m,std})} = \frac{\sum_m(\hat{b}_{m,std})(y_m - \bar{y})}{m-1} \tag{C.4}$$

$$= \frac{\left[\sum_m[\hat{\beta}_m - \overline{\hat{\beta}_m}]/\sqrt{\text{Var}(\hat{\beta}_m)}\right](y_m - \bar{y})}{m-1} = \frac{1}{\sqrt{\text{Var}(\hat{\beta}_m)}} \frac{\sum_{m=1}(\hat{\beta}_m - \overline{\hat{\beta}_m})(y_m - \bar{y})}{m-1} \tag{C.5}$$

where I use in the first line that $\text{Var}(\hat{b}_{m,std}) = 1$ and $\overline{\hat{b}_{m,std}} = 0$ because $\hat{b}_{m,std}$ is a standardized variable. In the second line I substitute in $\sum_m[\hat{\beta}_m - \overline{\hat{\beta}_m}]/\sqrt{\text{Var}(\hat{\beta}_m)}$ for $\hat{b}_{m,std}$.

For each major m the term $(y_m - \bar{y})$ is equivalent to $\hat{\beta}_m - \overline{\hat{\beta}_m}$. To see this first derive $\overline{\hat{\beta}_m} = \big(\hat{\beta}_1 + \hat{\beta}_2 + \cdots + \hat{\beta}_o + \hat{\beta}_m\big)/m = \big(\sum_{m \neq o} \bar{y}_m - \bar{y}_0(m-1)\big)/m$. Next, add and subtract $\overline{y_o}$ and use the law of iterated expectations ($E[Y] = \sum_i E[Y|A_i]P(A_i)$) to yield $E[y] = \sum_m E[y|m]p(m) = \sum_m \bar{y}_m \frac{n_m}{m} = \bar{y}_o \frac{n_o}{n} + \sum_{m \neq o} \bar{y}_m \frac{n_m}{m}$ where the final result follows since it is assumed that there is one observation per major so that $n_m = 1$ and $n = m$. Rearranging and simplifying terms yields: $\bar{y}_m - \bar{y} = (\bar{y}_m - \bar{y}_o) - \big(\sum_{m \neq o} \bar{y}_m - \bar{y}_o(m-1)\big)/m = \hat{\beta}_m - \overline{\hat{\beta}_m}$. Substituting into $\hat{a_m}$ yields:

$$\hat{a}_m = \frac{1}{\sqrt{\mathrm{Var}(\hat{\beta}_m)}} \frac{\sum_{m=1}(\hat{\beta}_m - \overline{\hat{\beta}_m})(y_m - \bar{y})}{m-1} \tag{C.6}$$

$$= \frac{1}{\sqrt{\mathrm{Var}(\hat{\beta}_m)}} \frac{\sum_{m=1}\left[\hat{\beta}_m - \overline{\hat{\beta}_m}\right]^2}{m-1} = \frac{\mathrm{Var}(\hat{\beta}_m)}{\sqrt{\mathrm{Var}(\hat{\beta}_m)}} = \sqrt{\mathrm{Var}(\hat{\beta}_m)} \tag{C.7}$$

This shows that $\hat{a}_1 = \sigma(\hat{\beta}_m)$: the coefficient on $\hat{b}_{m,std}$ conditional on X_{imt} – $\hat{a}_1$ – is exactly the standard deviation of the major fixed effects from a regression with m major fixed effects also conditional on X_{imt} – $\sigma(\hat{\beta}_m)$.

C.3.2 Equivalence of b_x in the two base models

In what follows I show that the coefficients on X_{imt} will be equivalent in Equations (C.1) and (C.2): $\hat{\beta}_x = \hat{\alpha}_x$. Note that these results will only follow if exactly the vector X_{imt} is equivalent in the two specifications.

A short proof of this result can be demonstrated using the omitted variable bias (OVB) formula. Suppose that the *naive* regression is $y = d_0 + d_x X + \epsilon$ which does not include either $b_{m,std}$ or the major dummies D_m and the *imt* subscript is omitted for notational clarity. Suppose the *true* models are Equations (C.1) and (C.2). The OVB formula yields two different expressions for the coefficient on X, one from each equation: $E[d_x] = \beta_x + \sum_m \beta_m cov(\mathbb{1}_m, x)$ and $E[d_x] = \alpha_x + \alpha_m cov(\hat{b}_{m,std}, X)$. Implicitly both expressions have $\mathrm{Var}(\hat{b}_{m,std})$ in the denominator as it is equal to one (since $b_{m,std}$ is a standardized variable with mean 0 and standard deviation 1). The coefficient on X is equivalent in both specifications if $\hat{\beta}_x = \hat{\alpha}_x$ which is equivalent to showing that $\sum_m \beta_m cov(\mathbb{1}_m, X) = \alpha_m cov(\hat{b}_{m,std} X)$ since $E[d_x]$ is the same in both equations. The OVB and some algebraic manipulation show that both expressions are equivalent to: $E[d_x] - \Bigg(\sum_i^n \left[(\sum_m \hat{\beta}_m \mathbb{1}_{mi} - \overline{\hat{\beta}_m})(X_i - \bar{X})\right]\Bigg)/(n-1)$.

Substituting the definition of $\hat{b}_{m,std}$ into $\alpha_m cov(\hat{b}_{m,std}, X)$ and noting that the mutual exclusivity of majors across a given observations implies that for observation i $\hat{\beta}_{m,i} =$

$\sum_m \hat{\beta}_m \mathbb{1}_{m,i}$ yields

$$\begin{aligned}
\alpha_m cov(\hat{b}_{m,std}, X) &= \alpha_m \frac{\sum_i^n (\hat{b}_{i,m,std} - \overline{\hat{b}_{m,std}})(X_i - \bar{X})}{n-1} \\
&= \sqrt{\mathrm{Var}(\hat{b}_{m,std})} \frac{\sum_i^n \left[(\hat{\beta}_{m,i} - \overline{\hat{\beta}_m})/\sqrt{\mathrm{Var}(\hat{b}_{m,std})}\right](X_i - \bar{X})}{n-1} \\
&= \frac{(\sum_m \hat{\beta}_m \mathbb{1}_{m,i} - \overline{\hat{\beta}_m})(X_1 - \bar{X}) + \cdots + (\sum_m \hat{\beta}_m \mathbb{1}_{m,i} - \overline{\hat{\beta}_m})(X_n - \bar{X})}{n-1}.
\end{aligned}$$

For the regression with m variables $\mathbb{1}_m$, the linearity of covariance means that:

$$\begin{aligned}
&\sum_m \hat{\beta}_m cov(hat\beta_m \mathbb{1}_{m=i}, X) = cov(X, \sum_m \hat{b}_m \mathbb{1}_m) \\
&= \frac{\sum_i^n (X_i - \bar{X})\left[(\hat{\beta}_1 \mathbb{1}_{m=1} + \hat{\beta}_2 \mathbb{1}_{m=2} + \cdots + \hat{\beta}_m \mathbb{1}_{m=i}) - \overline{\hat{\beta}_1 \mathbb{1}_{m=1} + \hat{\beta}_2 \mathbb{1}_{m=2} + \cdots + \hat{\beta}_m \mathbb{1}_{m=i}}\right]}{n-1}
\end{aligned}$$

The mutual exclusivity of majors for a given observation also implies $\hat{\beta}_1 \mathbb{1}_{1i} + \cdots + \hat{\beta}_m \mathbb{1}_{mi} = \sum_m \hat{\beta}_m \mathbb{1}_{mi}$ for each i. If we assume for expositional ease that that there is exactly one observation per major so that $n = m$ and $n_m = 1$) then for each i we also have that $\overline{\hat{\beta}_1 \mathbb{1}_{1i} + \cdots + \hat{\beta}_m \mathbb{1}_{mi}} = \sum_i (\hat{\beta}_1 \mathbb{1}_{1i} + \cdots + \hat{\beta}_m \mathbb{1}_{mi}) = \frac{\sum_m \hat{\beta}_m}{m} = \bar{\hat{\beta}}$. The desired result comes from direct substitution of these expressions.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock (2009): "5. The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in *Producer dynamics*, University of Chicago Press, 149–234.

Acemoglu, D. and D. Autor (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016a): "The analysis of field choice in college and graduate school: Determinants and wage effects," in *Handbook of the Economics of Education*, Elsevier, vol. 5, 305–396.

Altonji, J. G., E. Blom, and C. Meghir (2012): "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annu. Rev. Econ.*, 4, 185–223.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2014): "Trends in earnings differentials across college majors and the changing task composition of jobs," *American Economic Review*, 104, 387–93.

——— (2016b): "Cashier or consultant? Entry labor market conditions, field of study, and career success," *Journal of Labor Economics*, 34, S361–S401.

Altonji, J. G. and L. Zhong (2021): "The labor market returns to advanced degrees," *Journal of Labor Economics*, 39, 303–360.

Andrews, R., S. Imberman, M. Lovenheim, and K. Stange (2019): "The Returns to College Major Choice: Mean Effects, Career Trajectories, and Earnings Volatility," .

Andrews, R. J., S. A. Imberman, and M. F. Lovenheim (2017): "Risky business? The effect of majoring in business on earnings and educational attainment," Tech. rep., National Bureau of Economic Research.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020): "The evolution of work in the United States," *American Economic Journal: Applied Economics*, 12, 1–34.

Autor, D. H. and M. J. Handel (2013): "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 31, S59–S96.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 118, 1279–1333.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2016): "It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States," *Journal of Labor Economics*, 34, S67–S97.

Barth, E., J. Davis, and R. B. Freeman (2018): "Augmenting the human capital earnings equation with measures of where people work," *Journal of Labor Economics*, 36, S71–S97.

Barth, E., S. P. Kerr, and C. Olivetti (2021): "The dynamics of gender earnings differentials: evidence from establishment data," *European Economic Review*, 134, 103713.

Black, D., N. Kolesnikova, and L. Taylor (2009): "Earnings functions when wages and prices vary by location," *Journal of labor economics*, 27, 21–47.

Bleemer, Z. and A. Mehta (2022): "Will studying economics make you rich? A regression discontinuity analysis of the returns to college major," *American Economic Journal: Applied Economics*, 14, 1–22.

Blinder, A. S. (1973): "Wage discrimination: reduced form and structural estimates," *Journal of Human resources*, 436–455.

Blom, E., B. C. Cadena, and B. J. Keys (2021): "Investment over the business cycle: Insights from college major choice," *Journal of Labor Economics*, 39, 000–000.

Busso, M., J. S. Muñoz, and S. Montaño (2020): "Unbundling returns to degrees and skills: Evidence from postsecondary education in Colombia," Tech. rep., IDB Working Paper Series.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 128, 967–1015.

Cardoso, A. R., P. Guimaraes, P. Portugal, and H. Reis (2018): "The Returns to Schooling Unveiled," Tech. rep., IZA Discussion Papers.

Deming, D. and L. B. Kahn (2018): "Skill requirements across firms and labor markets: Evidence from job postings for professionals," *Journal of Labor Economics*, 36, S337–S369.

Deming, D. J. (2017): "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 132, 1593–1640.

Deming, D. J. and K. Noray (2020): "Earnings dynamics, changing job skills, and STEM careers," *The Quarterly Journal of Economics*, 135, 1965–2005.

Elder, T. E., J. H. Goddeeris, and S. J. Haider (2010): "Unexplained gaps and Oaxaca–Blinder decompositions," *Labour Economics*, 17, 284–290.

Engbom, N. (2022): "Labor market fluidity and human capital accumulation," Tech. rep., National Bureau of Economic Research.

Engbom, N. and C. Moser (2017): "Returns to education through access to higher-paying firms: Evidence from US matched employer-employee data," *American Economic Review*, 107, 374–78.

Foote, A. and K. Stange (2019): "Attrition from Administrative Data: Problems and Solutions with an Application to Higher Education," .

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition methods in economics," in *Handbook of labor economics*, Elsevier, vol. 4, 1–102.

Gelbach, J. B. (2016): "When do covariates matter? And which ones, and how much?" *Journal of Labor Economics*, 34, 509–543.

Gibbons, R. and M. Waldman (1999): "A theory of wage and promotion dynamics inside firms," *The Quarterly Journal of Economics*, 114, 1321–1358.

——— (2006): "Enriching a theory of wage and promotion dynamics inside firms," *Journal of Labor Economics*, 24, 59–107.

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

Goldin, C. D. and L. F. Katz (2009): *The race between education and technology*, harvard university press.

Haltiwanger, J., H. Hyatt, and E. McEntarfer (2018a): "Who moves up the job ladder?" *Journal of Labor Economics*, 36, S301–S336.

Haltiwanger, J. C., H. R. Hyatt, L. B. Kahn, and E. McEntarfer (2018b): "Cyclical job ladders by firm size and firm wage," *American Economic Journal: Macroeconomics*, 10, 52–85.

Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2017): "General education, vocational education, and labor-market outcomes over the lifecycle," *Journal of human resources*, 52, 48–87.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013): "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Tech. rep., National Bureau of Economic Research.

Hemelt, S. W., B. Hershbein, S. M. Martin, and K. M. Stange (2021a): "College Majors and Skills: Evidence from the Universe of Online Job Ads," Tech. rep., National Bureau of Economic Research.

Hemelt, S. W., K. M. Stange, F. Furquim, A. Simon, and J. E. Sawyer (2021b): "Why is math cheaper than English? Understanding cost differences in higher education," *Journal of Labor Economics*, 39, 397–435.

Hersch, J. (2013): "Opting out among women with elite education," *Review of Economics of the Household*, 11, 469–506.

Hershbein, B. and L. B. Kahn (2018): "Do recessions accelerate routine-biased technological change? Evidence from vacancy postings," *American Economic Review*, 108, 1737–72.

Hershbein, B. and M. Kearney (2014): "Major decisions: What graduates earn over their lifetimes," *Washington: Hamilton Project*.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019): "The allocation of talent and us economic growth," *Econometrica*, 87, 1439–1474.

Huneeus, C., F. Huneeus, C. Miller, C. Neilson, and S. Zimmerman (2021a): "Firm Sorting, Field of Study, and the Gender Earnings Gap," *unpublished manuscript*.

Huneeus, F., C. Miller, C. Neilson, and S. Zimmerman (2021b): "Firm Sorting, College Major, and the Gender Earnings Gap," .

Jacob, B. and L. Lefgren (2005): "Principals as agents: Subjective performance measurement in education," .

Jacob, B. A. and L. Lefgren (2008): "Can principals identify effective teachers? Evidence on subjective performance evaluation in education," *Journal of labor Economics*, 26, 101–136.

Jann, B. (2008): "The Blinder–Oaxaca decomposition for linear regression models," *The Stata Journal*, 8, 453–479.

Jovanovic, B. (1979): "Job matching and the theory of turnover," *Journal of political economy*, 87, 972–990.

Kinsler, J. and R. Pavan (2015): "The specificity of general human capital: Evidence from college major choice," *Journal of Labor Economics*, 33, 933–972.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 131, 1057–1111.

Lang, K. and M. Manove (2011): "Education and labor market discrimination," *American Economic Review*, 101, 1467–96.

Leighton, M. and J. D. Speer (2020): "Labor market returns to college major specificity," *European Economic Review*, 128, 103489.

Lemieux, T. (2014): "Occupations, fields of study and returns to education," *Canadian Journal of Economics/Revue canadienne d'économique*, 47, 1047–1077.

Li, X., S. Linde, and H. Shimao (2021): "Major Complexity Index and College Skill Production," *Available at SSRN 3791651*.

Long, M. C., D. Goldhaber, and N. Huntington-Klein (2015): "Do completed college majors respond to changes in wages?" *Economics of Education Review*, 49, 1–14.

Lovenheim, M. F. and J. Smith (2022): "Returns to Different Postsecondary Investments: Institution Type, Academic Programs, and Credentials," Tech. rep., National Bureau of Economic Research.

Malamud, O. (2010): "Breadth versus depth: the timing of specialization in higher education," *Labour*, 24, 359–390.

Martin, S. (2022): "Job Search and Earnings Growth: General and Specific Majors," Tech. rep., Unpublished Manuscript.

Minaya, V. and J. Scott-Clayton (2019): "Labor Market Outcomes and Postsecondary Accountability: Are Imperfect Metrics Better Than None?" *Productivity in Higher Education*, 67.

Neal, D. (1999): "The complexity of job mobility among young men," *Journal of labor Economics*, 17, 237–261.

Neal, D. A. and W. R. Johnson (1996): "The role of premarket factors in black-white wage differences," *Journal of political Economy*, 104, 869–895.

Nordin, M., I. Persson, and D.-O. Rooth (2010): "Education–occupation mismatch: Is there an income penalty?" *Economics of education review*, 29, 1047–1059.

Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets," *International economic review*, 693–709.

Oreopoulos, P., T. Von Wachter, and A. Heisz (2012): "The short-and long-term career effects of graduating in a recession," *American Economic Journal: Applied Economics*, 4, 1–29.

Ost, B., W. Pan, and D. A. Webber (2019): "Institution, Major, and Firm-Specific Premia: Evidence from Administrative Data," .

Paglin, M. and A. M. Rufolo (1990): "Heterogeneous human capital, occupational choice, and male-female earnings differences," *Journal of labor economics*, 8, 123–144.

Pavan, R. (2011): "Career choice and wage growth," *Journal of Labor Economics*, 29, 549–587.

Ransom, M. R. and A. Phipps (2017): "The Changing Occupational Distribution by College Major," in *Skill Mismatch in Labor Markets*, Emerald Publishing Limited.

Ransom, T. (2021): "Selective migration, occupational choice, and the wage returns to college majors," *Annals of Economics and Statistics*, 45–110.

Robst, J. (2007): "Education and job match: The relatedness of college major and work," *Economics of Education Review*, 26, 397–407.

Ruggles, S., S. Flood, R. Goeken, J. Grover, and E. Meyer (2020): "Jose Pacas, and Matthew Sobek,"IPUMS USA: Version 10.0 [dataset]," 2021," *As of October*, 23.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming up inequality," *The Quarterly journal of economics*, 134, 1–50.

Spitz-Oener, A. (2006): "Technical change, job tasks, and rising educational demands: Looking outside the wage structure," *Journal of labor economics*, 24, 235–270.

Syverson, C. (2011): "What determines productivity?" *Journal of Economic literature*, 49, 326–65.

Topel, R. H. and M. P. Ward (1992): "Job mobility and the careers of young men," *The Quarterly Journal of Economics*, 107, 439–479.

Webber, D. A. (2014): "The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors," *Labour economics*, 28, 14–23.

Wiswall, M. and B. Zafar (2015): "How do college students respond to public information about earnings?" *Journal of Human Capital*, 9, 117–169.

Wozniak, A. (2010): "Are college graduates more responsive to distant labor market opportunities?" *Journal of Human Resources*, 45, 944–970.

Zafar, B. (2013): "College major choice and the gender gap," *Journal of Human Resources*, 48, 545–595.