Essays in Labor Economics

by

Billy Haoyang Huang

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
2025

Doctoral Committee:

Professor John Bound, Co-Chair
Professor Jagadeesh Sivadasan, Co-Chair
Professor Sebastian Sotelo
Professor Kevin Stange

Billy Haoyang Huang
bhhuang@umich.edu
ORCID iD: 0009-0000-3481-1758

© Billy Haoyang Huang 2025

DEDICATION

I dedicate this dissertation to Linde, Adrielle, and Josephine, and to Shewen and Lucy.

ACKNOWLEDGEMENTS

Completing this dissertation[1] would not have been possible without the support of many people in my community of advisors, faculty, classmates, friends, and family.

I am deeply grateful to my committee members John Bound, Jagadeesh Sivadasan, Sebastian Sotelo, and Kevin Stange for their advice, support, wisdom, and unflagging encouragement throughout graduate school. Their intellectual engagement, invaluable feedback, and selfless, time-consuming investment in me have indelibly shaped my growth as an economist.

I also thank the faculty of the Michigan Economics department and the Gerald R. Ford School for their feedback and support, including Charlie Brown, Kathryn Dominguez, Mitchell Dudley, John Leahy, and Mel Stephens. I also thank numerous Michigan alumni for their help, including Henry Hyatt, Parag Mahajan, and Bryan Stuart.

My graduate school classmates have been an indispensable part of making it through graduate school, both professionally and personally. Thank you to George Cui, Andrea Foschi, Chris Hollrah, Emily Horton, Emir Murathanoglu, Hanna Onyshchenko, Nikhil Rao, Brock Rowberry, Taylor Thrasher, Levi Turner, Rachel Yang, and too many more Michigan graduate students and alumni to name. Special thanks to Heesung Kim and Stephen Tino for their unceasing and patient friendship, encouragement and moral support over many long conversations.

Finally, I am deeply blessed to have my family for their unconditional love and support. Kasey and Michael, thank you for always believing in and rooting for me. Shewen and Lucy, I can never repay your sacrifices and I am grateful to be your son. Linde, Adrielle, and Josephine—words cannot adequately capture the depth of self-giving love that you've poured out on me throughout this journey. You are my inspiration, my light and peace.

[1]Chapters 1 and 3 use 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. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2103. (CBDRB-FY24-P2103-R11314, CBDRB-FY24-P2103-R11373, CBDRB-FY24-P2103-R11746, CBDRB-FY25-P2103-R12480, CBDRB-FY25-P2103-R12496, and CBDRB-FY25-P2103-R12534.)

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vi

LIST OF TABLES viii

LIST OF APPENDICES x

ABSTRACT xi

CHAPTER

1 Coworker Networks and Internal Migration in the United States: Evidence from US Administrative Data 1

1.1 Introduction 1
1.2 Literature 3
1.3 Data 4
1.3.1 Data Description 4
1.3.2 Sample Construction 5
1.4 Empirical Stategy 7
1.4.1 Main specification 7
1.4.2 Discussion and identification assumptions 8
1.4.3 Coworker Tenure 9
1.4.4 Heterogeneity 9
1.5 Results 10
1.5.1 Main results 10
1.5.2 Destination firm subsamples 10
1.5.3 Distance 12
1.5.4 Coworker Tenure 12
1.5.5 Heterogeneity 12
1.6 Robustness checks and alternative samples 13
1.6.1 Excluding disappearing, shrinking, and growing firms 13
1.6.2 Multi-establishment firms 13
1.6.3 Movers and nonmovers 13
1.7 Conclusion 14

1.8 Figures . . . 15
1.9 Tables . . . 16

2 Labor Specialization and Search Frictions . . . 26

2.1 Introduction . . . 26
2.2 Literature . . . 28
2.3 Labor Specialization and Generality . . . 29
2.3.1 Data . . . 29
2.3.2 Measurement . . . 30
2.4 Matching Efficiency and Labor Specialization . . . 33
2.4.1 Matching Function Modeling and Estimation . . . 33
2.4.2 Unemployment and New Hires from CPS Data . . . 35
2.4.3 Results . . . 36
2.5 Conclusion . . . 38
2.6 Tables . . . 41
2.7 Figures . . . 52

3 Residential and Employment Locations in the Post-Pandemic US: Evidence from US Administrative Data . . . 57

3.1 Introduction and motivation . . . 57
3.2 Literature . . . 58
3.3 Data . . . 59
3.4 Results . . . 60
3.4.1 Increasing post-pandemic home-work distances . . . 61
3.4.2 Hiring cohort . . . 62
3.4.3 Industry . . . 62
3.4.4 Gender . . . 62
3.5 Conclusion . . . 63
3.6 Figures . . . 64

APPENDICES . . . 71

BIBLIOGRAPHY . . . 101

LIST OF FIGURES

FIGURE

1.1 30 LEHD states available 15

2.1 CDF of Skills 52
2.2 Histogram of Job Generality: Main Generality Measure 53
2.3 Histogram of Job Generality: Alternative Generality Measure 54
2.4 Histogram of Number of Skills 55
2.5 Scatterplot of Skill Generality vs. Number of Skills 56

3.1 Mean Distances 64
3.2 Median Distances 65
3.3 Percentage of Employees Living 50+ miles from Work 66
3.4 Average Distance, by hire cohort 67
3.5 Percentage of Employees Living 50+ miles from Work, by hire cohort 68
3.6 Percentage of Employees Living 50+ miles from Work, by Industry 69
3.7 Percentage of Employees Living 50+ miles from Work, by Gender 70

B.1 Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation . . 77
B.2 Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation 78
B.3 Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation 80
B.4 Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation 81
B.5 Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation . . 82
B.6 Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation 83
B.7 Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation 84
B.8 Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation 85
B.9 Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation . . 86
B.10 Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation 87

B.11 Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation 88
B.12 Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation 89
B.13 Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation . . 90
B.14 Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation 91
B.15 Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation 92
B.16 Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation 93

C.1 Mean Distances, quarterly 94
C.2 Median Distances, quarterly 95
C.3 Percentage of Employees Working 50+ miles from Home, quarterly 96
C.4 Average Distance, by hire cohort, quarterly 97
C.5 Percentage of Employees Living 50+ miles from Work, by hire cohort, quarterly 98
C.6 Percentage of Employees Living 50+ miles from Work, by Industry, quarterly . 99
C.7 Percentage of Employees Living 50+ miles from Work, by Gender, quarterly . . 100

LIST OF TABLES

TABLE

1.1 Summary statistics: Main Sample 16
1.2 Summary statistics: 30-state sample 16
1.3 Summary statistics: Employer Change sample 17
1.4 Main Sample: Annual Earnings by Percentile 17
1.5 Effect of *coworker* on *comove*: Main Sample 18
1.6 Effect of *coworker* on *comove*: 30 State, Employer Change, and Different Destination Firm Subsamples 18
1.7 Effect of *coworker* on *comove*: Distance subsamples 19
1.8 Effect of *coworker* shared tenure on *comove* 19
1.9 Effect of *coworker* on *comove*: Heterogeneous specification, race and ethnicity . 20
1.10 Effect of *coworker* on *comove*: Heterogeneous specification, gender 21
1.11 Effect of *coworker* on *comove*: Heterogeneous specification, place of birth . . . 21
1.12 Effect of *coworker* on *comove*: Heterogeneous specification, age 22
1.13 Effect of *coworker* on *comove*: Heterogeneous specification, pre-migration wage 23
1.14 Effect of *coworker* on *comove*: Excluding firms that shrink or grow by more than 30% 24
1.15 Effect of *coworker* on *comove*: Movers and non-movers 25

2.1 25 Most Common Skills and their Frequency 41
2.2 All Occupations Ranked from Most to Least General, SOC-2 Level: Main Generality Measure 42
2.3 All Occupations Ranked from Most to Least General, SOC-2 Level: Alternative Generality Measure 43
2.4 Top 25 Most General Occupations, SOC-6 Level: Main Generality Measure . . 44
2.5 Top 25 Most General Occupations, SOC-6 Level: Alternative Generality Measure 45
2.6 Top 25 Least General Occupations, SOC-6 Level: Main Generality Measure . . 46
2.7 Top 25 Least General Occupations, SOC-6 Level: Alternative Generality Measure 47
2.8 Matching Function Estimation, SOC-2: Contemporaneous vacancies and unemployment 48
2.9 Matching Function Estimation, SOC-2: Lagged vacancies and unemployment . 49
2.10 Matching Efficiency and Occupational Generality, Contemporaneous Matching Function Estimation 50
2.11 Matching Efficiency and Occupational Generality, Lagged Matching Function Estimation 50
2.12 Vacancy and Generality Variation Across MSA Types, 2010-2019 51

A.1 Effect of *coworker* on *comove*: Heterogeneous specification 73
A.2 Effect of *coworker* on *comove*: Single and multi-establishment firms 74
A.3 Effect of *coworker* on *comove*: Movers and nonmovers, manufacturing 75

B.1 Vacancy and Generality Variation Across States, 2010–2019 79

LIST OF APPENDICES

A Appendix to Chapter 1 . . . . . **71**
Data Construction 71
Appendix Tables 73
B Appendix to Chapter 2 . . . . . **76**
Appendix Tables and Figures 76
C Appendix to Chapter 3 . . . . . **94**
Appendix Figures - Quarterly 94

ABSTRACT

This dissertation studies how worker mobility and human capital are mediated by firms. Chapter I studies how coworker networks influence the internal migration of workers within the U.S. Chapter II studies how worker specialization affects job search. Chapter III studies how patterns in U.S. residential and workplace locations have shifted since the 2020 global pandemic.

In Chapter I, using administrative Census data, I present new empirical evidence of the effects of coworker networks on the location choice of migrants, consistent with information provision about opportunities in other local labor markets. Workers who migrate are significantly more likely to choose the same destination as that of their coworkers, than the destinations of non-coworkers from the same industry and location. A set of additional results suggest a role for information provision. Coworkers who have worked together longer have a greater likelihood of migrating to the same locations. Moves to the same destination firms account for only a small portion of the network effect, suggesting that direct job referrals play only a small role in facilitating such migration. I estimate stronger network effects among workers who move further distances, are older, and have higher earnings. Excluding moves associated with relocation within the same firm, and excluding establishment closures, do not significantly alter the results, consistent with a primary role of information provision about destination markets through coworkers.

In Chapter II, I examine the relationship between labor specialization and labor market search frictions. Using online job vacancy data, I construct occupation-level measures of specialization based on the rarity of skills associated with job postings. I also combine vacancy counts with hiring flows and unemployment from the Current Population Survey to estimate matching efficiency at multiple occupational aggregation levels. I find evidence that more specialized occupations are associated with lower matching efficiency, implying greater frictional unemployment. In standard macroeconomic labor search models, matching efficiency is typically treated as exogenous; my results suggest it may be systematically related to the economy's occupational composition. This tradeoff between the productivity benefits of specialization and the slower reallocation of labor suggests profound macroeconomic implications: it can affect aggregate unemployment, the position of the Beveridge curve, and

the speed of labor market adjustment after shocks.

In Chapter III, using data on worker residential location from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD), I study how the distance between U.S. workers' residential and employment locations has changed since the 2020 global pandemic. I find that mean and median work-home distances have increased since 2020 Q2. The increase in distance is largely driven by workers who have been hired since 2020 Q2, as opposed to workers hired before then, suggesting that the geographic labor market in which employers are searching has expanded as a result of the pandemic. However, from 2020 Q2 to 2022, the mean distance declines for those new hires, suggesting that an initial increase in distance has tempered over time, consistent with corporate return-to-office policies. Additionally, the increase in distance is driven primarily by jobs in the information technology industry, and to a lesser extent by jobs in the professional services, finance, and administration industries.

CHAPTER 1

Coworker Networks and Internal Migration in the United States: Evidence from US Administrative Data

1.1 Introduction

Internal migration is crucial for economic growth and efficiency, as it enables workers to relocate to areas optimal for their welfare and allows regions to adjust to local economic shocks. Understanding the determinants of and barriers to internal migration is essential for understanding how labor markets equilibrate following such shocks, for developing policies that influence regional outcomes, and for better matching of firms and workers across space.

While a number of variables influence migration, job-related factors are cited most frequently in survey data as motivation for long distance moves—for example, accounting for about half of inter-state moves in the U.S. (Jia et al. 2023). Yet the existence of large and growing regional income differences raises the question of whether or not such migration occurs at ideal rates (Nunn et al. 2018; Gaubert et al. 2021). Uncertainty about wages and amenities, and the cost of obtaining information about opportunities in other locations, represent important frictions to the equilibration of local labor markets and to potentially welfare- and efficiency-enhancing matches.

Moreover, social networks have been shown to be important in labor market information transmission (Ioannides and Loury 2004). However, relative to other types of social networks (such as those centered around family, neighborhoods, ethnic groups, etc.), coworker networks may be a particularly salient source of information relevant to labor market-motivated migration. Highly skilled workers may be even more likely to benefit from job-related information relayed through professional networks, as such information is less likely to come through casual social channels.

While the role of firm-centered networks in facilitating geographic mobility is not well

understood, it is likely to be significant, particularly for longer-distance moves across local labor markets. Given that job-related factors are the primary drivers of internal migration, access to information about job opportunities in potential destinations is crucial. Coworkers, with their knowledge of specific skills and potential matches, are a natural source of such information. Moreover, as coworkers often interact socially, they can also provide insights into the amenities and lifestyle factors of destination locations, making them a valuable resource for individuals considering geographic mobility.

This paper provides novel evidence on the importance of workplace networks in facilitating migration across labor markets. My findings suggest that workplace networks transmit information about job opportunities or amenities in distant local labor markets. To show this, I use data on residential and employment locations (at the commuting zone level), individual demographics, employer characteristics, and coworker relationships from the US Census Bureau's Longitudinal Employer-Household Dynamics (LEHD). Using these data, I implement a dyadic regression design on pairs of individuals[1]. The design detects network effects by comparing whether individuals who work at the same firm are more likely to move to the same destinations than individuals who work at different firms in the same industry and location.

My main finding is that pairs of workers who work at the same firm are 1.6 percentage points more likely to migrate, five years later, to the same commuting zones as workers who are not coworkers but work in the same industry. This estimate corresponds to 37% of the baseline propensity of 4.3 percentage points for non-coworker pairs from the same industry and origin to move to the same destination.

The key identification assumption in this setting is that there are no unobserved firm-level factors, within industry and location, that affect the propensity to migrate to a particular location. Thus, identification threats include factors that disproportionately influence workers from a particular firm to migrate to certain destinations that do *not* affect workers from other firms in the same industry and location. I address several plausible alternative explanations, such as migration corresponding to employment changes across establishments but within the same larger firm, or migration induced by establishment-specific closures and the resulting mass displacement, and find that the network effects remain robust. My main set of estimates focus on workers that move, though in a later exercise I expand the sample to also include non-migrants and still find significant network effects.

Next, in order to address concerns that my findings are being driven by within firm

[1]This type of design was first pioneered by Bayer et al. (2008), and similar empirical designs have been used to detect network effects in other settings, as in Schmutte (2015), Stuart and Taylor (2021), and Zhu (forthcoming).

moves and to disentangle the role of destination firms, I examine a subsample from my main analysis with complete data on worker employment in their destination. I find that only a small portion, about 15%, of the network effect comes from workers joining the same destination firms after they migrate. This suggests that direct referrals are not driving the network effects, and that workers are gaining information about opportunities in the location apart from the traditional referral channel.

Next, to bolster the interpretation that my results reflect information transmission within the network, I find that the network effects are stronger the longer coworkers work together: one additional quarter working together increases the probability that coworkers move to the same location by 2% of the baseline effect. Alternative explanations, such as firm-specific factors inducing correlated migration, are less plausible in light of this finding.

In studying heterogeneity in the estimated network effects, I find stronger effects among workers who are older, have higher baseline, pre-migration earnings, and who move further distances. This suggests that the types of workers for whom network effects are stronger tend to be more experienced and skilled, and that the importance of information grows across space.

In testing alternative possible explanations, I also find that excluding moves associated with relocation within the same firm, and excluding establishment closures, do not significantly alter the results, consistent with a primary role of information provision about destination markets through coworkers.

The remainder of the paper proceeds as follows. Section 3.2 discusses how this paper contributes to the existing literature. Section 3.3 describes the Census data utilized. Section 1.4 discusses the empirical strategy and identification assumptions. Section 3.4 discusses the main empirical results. Section 1.6 discusses several alternative explanations and robustness checks. Section 3.5 concludes. Additional details are provided in the appendix.

1.2 Literature

The role of social networks and information frictions is little studied in the internal migration literature. In the US, work by Carrington et al. (1996) and Stuart and Taylor (2021) find an important role for networks in facilitating migration in the Great Migration, one of the most important mass migration episodes in US history. In contrast, this paper is the first to study their importance in the context of the modern US economy.

Social networks more broadly have been shown to be important in migration decisions in the international migration context (Munshi 2020), for which there are important legal barriers. This paper, in contrast, focuses on the far more common phenomenon of internal

migration within the borders of a country or region, where workers are free to move wherever they choose, as soon as they are willing.

There is a small literature demonstrating the importance of information transmission for internal migration in developing countries based on randomized trials (Baseler 2023, Frohnweiler et al. 2022), structural frameworks (Porcher 2022), or particular migration episodes (Wilson 2022). This paper focuses on the developed US context, with a far larger set of labor markets, and for which the lessons from randomized trials may be less applicable.

Finally, this paper contributes to a growing literature on information exchange through social networks (Ioannides and Loury 2004), which consistently find that a wide range of network types impact labor market outcomes such as job search, wages, and unemployment duration, and entrepreneurship. Examples of such networks include those centered around residential neighbors (Bayer et al. 2008, Hellerstein et al. 2011 Schmutte 2015), coworkers (Cingano and Rosolia 2012, Caldwell 2019, Wallskog 2023), family (Kramarz and Skans 2014), and university classmates (Zimmerman 2019), or social contacts measured through phone records (Barwick et al. 2023, Blumenstock et al. forthcoming). There is increasing evidence that such networks operate heterogeneously as well, with referrals from business contacts being used more frequently than family referrals among more productive workers (Lester et al. 2023).

Much of this literature emphasizes job-to-job mobility, which most often occurs within the same local labor market, whereas this paper emphasizes longer distance migration across labor markets. In addition, much of this literature explicitly or implicitly focuses on the role that direct referrals play in job search and outcomes, whereas this paper additionally examines the role that broader information about distant markets play.

1.3 Data

1.3.1 Data Description

I primarily use the Longitudinal Employer Household Dynamics (LEHD) dataset in my analysis, which is restricted-use administrative linked employer-employee data from the U.S. Census Bureau. The LEHD covers 95% of private-sector employment in the US (Graham et al. 2022).

I use LEHD quarterly earnings data from 30 states [2], which collectively covered 51% of the US population in 2010 and has reasonable geographic representation of the US. For this

[2] AL, AZ, CO, CT, DC, DE, HI, ID, IL, KS, LA, ME, MD, NE, NV, NJ, NM, ND, OH, OK, OR, PA, SC, TN, TX, UT, VA, WA, WI, WY. See Figure 1.1 for a map.

set of states, I observe quarterly employment and earnings information, which means I can infer coworker relationships from workers that share a firm[3] at the same time.

In addition, the data contain annual individual information on residential location, at the Census block or tract level, for workers in all states, from 1999 to 2020. The data also contain information on worker demographics (including race, ethnicity, gender, age, and country of birth), and firm characteristics (such as firm location, 6-digit NAICS industry, payroll, and size). Additional data and sample restrictions are discussed in Appendix A.

1.3.2 Sample Construction

1.3.2.1 Main Sample

My main sample consists of workers aged 25-54 in 2014, who work in single-establishment[4] firms with firm sizes between 2 and 1000 workers, who reside in the 50 states and the District of Columbia, who change their commuting zone (CZ) of residence between 2014 and 2019, and who earn more than $15,000 annually across all jobs. I discuss each restriction below.

The age restriction focuses the analysis on prime-age workers who are less likely to make migration decisions for education or retirement-related reasons. The earnings restriction of $15,000 annually is roughly equivalent to the annual earnings of full-time work at the federal minimum wage of $7.25 an hour. This restriction is made to focus the analysis on full-time workers instead of part-time workers.

The single-establishment restriction is made because for multi-establishment firms, the exact establishment a worker works in is imputed in the data. Thus for the single-establishment sample, we have no uncertainty about which establishment a worker works at (and thus no uncertainty about which coworkers share their work location). However, in an alternative sample I relax this restriction to include multi-establishment firms as well, with robust results.

The firm size upper bound restriction is common in the coworker networks literature (Caldwell 2019), with the notion that in very large firms workers are unlikely to know each other. The lower bound is required for the worker to have any coworkers. Note that all of the firm characteristic restrictions are only for the origin firms, while the destination firms

[3]There are several notions of "firm" in the LEHD, but I use the term here primarily to refer to a business entity at the State Employer Identification Number (SEIN) level. Firms comprise establishments, but I primarily focus on single-establishment firms in my main analysis due to imputation reasons, as described below.

[4]To be more precise, single-establishment indicates firms that have only one establishment in a state, because this is a state employer identification number (SEIN). Thus, firms that span several states but have a single establishment in each state would be included. Firms that have more than one establishment within a state are excluded.

remain unrestricted.

Finally, I restrict the main sample to workers who change their commuting zone of residence between 2014 and 2019. Because the analysis conditions on workers who migrate, the relevant margin through which one should interpret the results is that of the choice of where to move, rather than the extensive margin of whether or not to move. Many push and pull factors may contribute to the decision to migrate or not; this restriction remains agnostic on those factors. This also partly addresses concerns about unobserved firm-level factors within industry and location affecting migration. I also relax this restriction in an alternative sample to include this extensive margin.

As workers sometimes have multiple jobs in one year, I keep a worker's highest paying job for 2014 as the worker's "main job" and drop other jobs. This main job defines the set of coworkers for a particular worker.

Then, in order to implement the empirical design, I construct a dyadic dataset consisting of pairs of all workers who work in the same 6-digit industry and whose employers are located in the same CZ. I make a few additional pair-level sample restrictions. To address the potential issue of correlated household-level comigration, such as married couples who happen to be coworkers making joint location decisions, I exclude pairs of workers who live at the same address (using an address identifier available for residential data after 2011).

I make a few further sample restrictions that do not change my final regressions results because they do not contribute to variation, but will inflate the sample size. I drop industry-location groups that have one or two workers, because these workers will be unpaired or will be excluded from the final regression as singletons, respectively. I also drop industry-location groups which contain only one firm, as workers in these firms will all be coworkers and there will be no comparison group, and thus no variation. This set of pairs constitutes my main sample.

Tables 1.1 and 1.4 show summary statistics and wages for the main sample.

1.3.2.2 Subsamples

Because the main sample described in section 1.3.2.1 includes workers who migrate to states where I have residential information but lack employment information, I further create two additional subsamples. First, I have a subsample that includes only those workers who work in one of the 30 states I have employment information upon migrating. I label this the "30-state sample" in the rest of the paper. Table 1.2 displays summary statistics for this sample.

Second, I create another subsample that further excludes workers who do not change employers upon moving. I label this the "employer change" sample in the rest of the paper.

This sample is motivated by three reasons. First, some workers migrate short distances across the border of a commuting zone, but keep the exact same job upon a move. These workers should be excluded from the main analysis. Second, we want to exclude workers that are migrating longer distances but moving within establishments of a larger firm potentially spanning across states, in order to rule out firm-level decisions that induce mass correlated migration but are not due to information transmission through the network[5]. Finally, we want to characterize the destination firms and industries. Table 1.3 displays summary statistics for this sample.

Because I have employment data for workers in 30 states, but residential information for workers in all 50 states, the subset of workers for whom I observe firm information in both the destination and origin is necessarily a subset of the main sample. Additionally, some workers may leave the labor force upon migrating, which reduces the sample of employer changers further. Consequently, the number of moving worker pairs in the employer change sub-sample is about 25 million, compared to the 349 million worker pairs in the full sample[6].

1.4 Empirical Stategy

1.4.1 Main specification

I adapt my empirical specification from Bayer et al. (2008), using a dyadic regression design estimated at the level of pairs of workers. Similar designs to detect network effects have been used in a variety of other settings, as in Schmutte (2015), Stuart and Taylor (2021), and Zhu (forthcoming). I construct an indicator variable $coworker_{ij}$ for whether or not a pair of workers i, j are coworkers in the same firm in 2014. I also construct an indicator variable $comove_{ij}$ for whether or not a pair of workers i, j moved to the same CZ by 2019.

Then I regress $comove_{ij}$ on $coworker_{ij}$, as follows:

$$comove_{ij} = \alpha_g + \beta_0 coworker_{ij} + \epsilon_{ij} \tag{1.1}$$

where ij are individuals, g is a 6-digit industry-CZ group, α_g represents the group fixed effect, i.e. the baseline probability moving to the same location d for individuals in the same group industry-CZ who are not coworkers.

The estimated β_0 represents the additional propensity of coworkers to move to the same location, on top of the baseline propensity for two workers in the same group (industry-CZ)

[5]However, as I discuss below in section 1.4.2, a robust conception of networks could encompass such within-firm moves.

[6]Note that the differences in magnitude are roughly squared when comparing pairs of workers.

who are not coworkers to move to the same location. Thus, dividing the β_0 by the mean of $comove_{ij}$, conditional on $coworker_{ij}$ being zero, gives the estimated propensity of interest.

The α_g fixed effect importantly ensures that the comparison is between pairs of coworkers and pairs of workers in the same industry and location. This accounts for a few obvious reasons why there might be correlated comovement of coworkers. First, industry-level shocks. Second, migration is more likely to nearby commuting zones.

Standard errors are clustered at the industry-CZ level, allowing for arbitrary correlation of observations within an industry-location group. This is an overly conservative assumption for inference, as we want to account for the obvious correlation among individuals who appear multiple times in the observations as members of pairs, but not necessarily all individuals within an industry-CZ. Future versions of this paper may use pairwise bootstrapping or multi-way clustering by individuals and industry-CZ to be more precise.

1.4.2 Discussion and identification assumptions

The empirical design essentially uses workers who are employed in the same industry and location as a comparison group to workers in the same firm. Thus, this controls for industry- and location-specific shocks that might disproportionately affect correlated migration choices.

The main identification assumption is that there is no correlation in unobserved factors affecting migration location choice among workers within the same industry and commuting zone. Thus, this strategy is valid to the extent that such shocks are industry and location specific, and not firm-specific within an industry-location. Potential threats to identification include unobserved factors that affect the migration propensity of workers from a particular firm but not the other firms in its industry and location. That is, we need to assume that workers do not sort into firms within an industry-CZ in a way that affects their subsequent residential mobility.

Another potential concern is the phenomenon of firms that are relocating workers between its own establishments, which I attempt to address using the employer change subsample, which excludes workers that relocate between establishments of firm, including larger, national firms that span state borders.

However, a robust conception of a coworker network can include within-firm moves, or even large groups of workers joining the same rival firm, as potential mechanisms for how the network operates. For instance, groups of workers who provide complementary skills and consequently increase their productivity as a team could experience reassignment by the firm to another establishment, or be poached as a group by a rival firm. Such phenomena would be consistent with a strong role for networks in facilitating migration, but

are difficult to disentangle from firm-level employment and relocation decisions that do not involve social network phenomena nor information transmission. Nevertheless, I test the importance of these phenomena under various sample restrictions in order to highlight the role of information transmission.

Another potential concern is firm demand for specific labor market skills. If few firms demand very particular sets of skills, and these firms are distributed across space, then correlated coworker migration could mostly reflect the fact that workers have the same few labor market opportunities, not network factors. I address this concern several ways. First, in the sample criteria, industry-location groups containing only a single firm are dropped, which helps exclude highly specialized firms. Second, I study the extent to which migrating workers join the same destination firms. If workers are not joining the same destination firms (and instead joining other firms in the destination location), this would inconsistent with both firm-initiated relocation and demand for specialized skills.

1.4.3 Coworker Tenure

I also study the extent to which the length of time a pair of coworkers has worked together affects their propensity to move to the same location, using an extended empirical specification which includes the number of quarters that a pair of workers has overlapped:

$$comove_{ij} = \alpha_g + \beta_0 coworker_{ij} + \beta_1 coworker_qtrs_{ij} + \epsilon_{ij}$$

where, $coworker_qtrs_{ij}$ captures the number of quarters that two coworkers have worked together.

1.4.4 Heterogeneity

To examine how the strength of the effects vary based on how similar workers are to each other (homophily), it is straightforward to extend equation (1) to allow for a variety of demographic characteristics X_{ij} as covariates, both in levels and interacted with $coworker_{ij}$. X_{ij} is a vector of indicators for whether the worker pair shares demographic characteristics, such as race, ethnicity, gender, country of birth, similar age, and similar initial wage buckets.

$$comove_{ij} = \alpha_g + (\beta_0 + \gamma_0' X_{ij}) coworker_{ij} + \gamma_1' X_{ij} + \epsilon_{ij} \tag{1.2}$$

The covariates term $\gamma_1' X_{ij}$ thus control for individual-specific reasons why two workers work in the same firm, accounting for sorting into the same firm. The coefficient on the interaction term γ_0 signifies which demographic characteristics shared by i and j increase

the propensity to move to the same location.

1.5 Results

1.5.1 Main results

The main results of regression (1) are in Table 1.5. The sample comprises about 1.6 million individuals, with roughly 350 million pairs of workers.

The estimated β shows that a pair of coworkers have a additional 1.6% probability of moving to the same destination commuting zone.

The baseline propensity of a pair of workers–who are not coworkers but work in the same industry and origin commuting zone–to move to the same destination commuting zone is 4.3%. This alone demonstrates a significant degree of correlated migration–if workers are randomly moving to the roughly 700 commuting zones in the country, the probability that any two migrate to the same one would be many orders of magnitude smaller.

To interpret the magnitude of the estimated β, we take its ratio with the baseline propensity to obtain an overall effect size of about 36.8%. Interpreted as a network effect, this magnitude is meaningfully large compared to other studies that use similar designs (33% in Bayer et al. (2008), 18% in Schmutte (2015)).

1.5.2 Destination firm subsamples

1.5.2.1 Subsample description

An important concern is whether or not such correlated movement is the result of relocation due to firm-related factors, such as an employer moving a group of workers from one regional office to another. Such co-migration would be due less to social factors, but to a decision by the firm[7].

However, because my sample only includes 30 states for earnings and employer-employee linkage, I cannot observe destination firms for everybody in my original sample, whereas I observe destination residential locations for all workers in the sample. If a worker moves to a state not in my sample, or drops out of the labor force entirely, then I will not observe their earnings nor employer in the destination. Thus, in order to both rule out concerns about employees migrating but moving jobs within the same larger firm, and to elucidate

[7]However, a sufficiently robust conception of networks could encompass even such firm-related decisions. For instance, groups of workers whose skills complement each other and become particularly productive working with each other might choose to migrate with each other, but such migration could also be induced by a firm that observes a particularly productive team.

mechanisms regarding destination firms, I use several subsamples of workers with non-missing destination firm information.

In Table 1.6, I display the results of regression (1) for three subsamples. Column (1) displays the results for the subsample of 30 states with non-missing destination information.

Column (2) takes the sample from column (1) but further eliminates pairs of workers who do not change jobs. This encompasses both workers who migrate and switch jobs within the same national firm, but also workers who do not change firms at all (even staying within the same establishment) while migrating (e.g., a short distance, across the border of their commuting zone).

Finally, column (3) takes the sample from column (2) but drops pairs in which both workers in the pair join the same destination firm. The underlying set of individuals remains the same, but certain combinations of workers are dropped, so the number of pairs is smaller compared to column (2).

Summary statistics for the sample in column (1) are in table 1.2, and summary statistics for the sample in columns (2) and (3) are in table 1.3.

1.5.2.2 Subsample results

In column (1), we see that the number of individuals is about a fourth of the main sample. Due to the fact that there are fewer commuting zones in this subsample of 30 states, the baseline propensity will rise mechanically: fewer possible destinations implies the probability of two workers joining the same destination will increase. The baseline propensity is 17%, while the estimated network effect is 4.0%, for an overall effect size of 23.4%.

In column (2), the baseline effect is 15.4%, and the estimated network effect is 3.1%, for an overall effect size of 19.9%. This suggests that workers staying with the same firm upon migration accounts for some of the network effect, but the majority of the effect is still present after accounting for this.

Finally, in column (3), the estimated coefficient is now 2.6%, for an overall effect size of 16.8%. This represents a roughly 15% decrease from column (2), suggesting that moves to the same destination firm account for a minority of the network effect. Workers joining the same destination firms is suggestive of, or a proxy for, direct job referrals. These results seems to suggest that workers are gaining information about the destination location itself through their coworkers, and not only through referrals through jobs.

1.5.3 Distance

Table 1.7 shows the results of regression (1) for various distance subsamples, in which only workers who move more than 50, 100, or 250 miles are included. These distances represent one or several hours of driving distance, meant to capture workers. This is an alternative method of excluding workers who are moving simply across the border of a CZ but are effectively not changing their local labor market. The benefit of this specification is that I can exclude workers who migrate just across commuting zone borders without necessarily needing information on destination firms, as I do in the previous section, which reduces the sample considerably.

For the subsample moving at least 50 miles, the network effect is 1.6% on a baseline of 3.3%, for an overall effect size of 47.0%. For the subsample moving at least 100 miles, the network effect is 1.7% on a baseline of 3.2%, for an overall effect size of 52.9%. For the subsample moving at least 250 miles, the network effect is 1.9% on a baseline of 3.0%, for an overall effect size of 66.2%.

Overall, the network effects are stronger than in the main sample and interestingly, appear to be stronger as the distance grows. Viewed in combination with the results from the previous section, this suggests that workers moving short distances but staying in the same firm is not a major concern.

1.5.4 Coworker Tenure

Table 1.8 shows the results of regression (2) using the main sample. The mean number of quarters that coworkers have worked together is 6.9%. The estimated effect on *coworker* is 1.0%, off of a baseline of 4.3%, for an overall effect size of 23.8%. The estimated effect on *coworker_qtrs* is .00083, which corresponds to a 1.9% overall effect size for each additional quarter. This shows that workers who work longer together are more likely to move to the same location, which is consistent with a social network effect and an information transmission mechanism.

1.5.5 Heterogeneity

Table A.1 shows the results of regression (3). I report the estimates on the interaction term with *coworker*. The results suggest that pairs in which both workers are older and have higher initial wages see strong network effects. The estimated magnitudes as a proportion of the estimated coefficient on *coworker* for both are large in magnitude.

Tables 1.9 through 1.13 show these results split up by demographic category, but demonstrate essentially the same results.

1.6 Robustness checks and alternative samples

1.6.1 Excluding disappearing, shrinking, and growing firms

A potential concern is that of firm closures causing groups of workers to comove to other locations for reasons other than the network. For example, displaced workers may move together not because of their network but because of a firm closure shock that differentially affects where workers choose to move. On the flip side, successful, fast-growing firms may see groups of workers being targeted for poaching by other firms, so such workers may migrate together as well. To account for these concerns, I estimate the effects on a sample that excludes firms that shrink or grow in size by more than 30% between 2014 and 2019 and find similar effect. 30% is the standard cutoff for measuring mass layoff events in the displaced worker literature (Flaaen et al. 2019). The results are in Table 1.14, with estimated network effects similar to those in the main results.

1.6.2 Multi-establishment firms

For firms with more than one establishment within a state, the exact establishment that a worker is employed in is imputed by the LEHD, so in my main sample I include only firms with one establishment in a state because there is no uncertainty about the employment location of the worker. Thus I analyze an alternative sample, roughly double in size, that includes such multi-establishment firms. The results are in Table A.2, with estimated network effects similar to those in the main results.

1.6.3 Movers and nonmovers

Finally, in Tables 1.15 and A.3, I include workers who do not move, and define the comove variable to equal 1 if both workers in a pair move to the same location, and 0 if workers are moving to different locations or one of the workers does not move at all. This increases the sample size considerably, so I use a 10% random sample in Table 1.15, and restrict the origin industries to manufacturing in Table A.3. Because most workers do not migrate, the estimated β and baseline propensities are very small. However, the estimated overall effect size is quite large, at 72.0% for the sample with all industries. This suggests that selecting on moving does not significantly bias the results.

1.7 Conclusion

In this research, I shed light on an understudied channel of information transmission facilitating migration, finding evidence for coworker network effects in migration across local labor markets in the US. Coworker pairs are more likely to migrate to the same location than pairs of workers who work in the same industry and location. Direct referrals, interpreted as moves to the same destination firms, account for only a minority of the effect. Coworker pairs who have worked together longer are also more likely to migrate to the same location as well, further supporting the notion of transmission of information across networks. I find that excluding moves associated with relocation within the same firm, and excluding establishment closures that result in mass layoff events, do not significantly alter the results. A better understanding of the mechanisms behind such migration is crucial for better designing policies that enhance labor market fluidity across space, and address the frictions that hinder migration and affect spatial inefficiency. While the type of information transmitted through the networks studied in this paper are particular and may be difficult to replicate, it is clear that policies that incentivize migration should account for potential spillover or multiplier effects that could affect the migration of additional workers connected through a network.

1.8 Figures

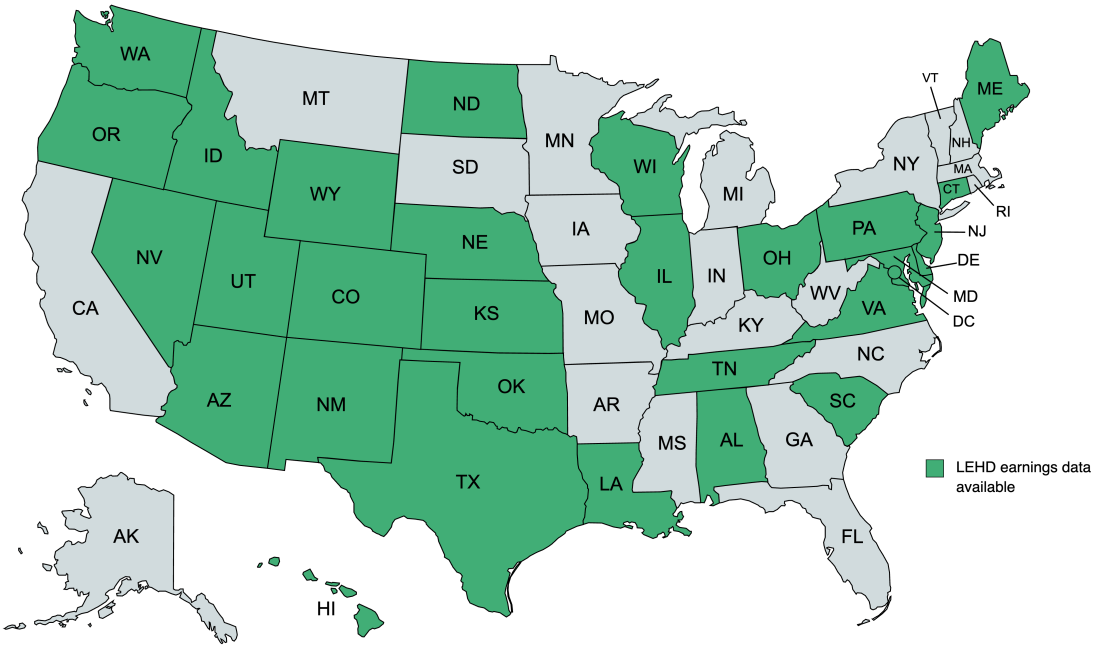


Figure 1.1: 30 LEHD states available

1.9 Tables

Table 1.1: Summary statistics: Main Sample

Variable	Value
Mean Age	35.03
Median Age	34
Male (%)	59.9
Native-born (%)	84.4
White (%)	79.8
Black (%)	9.6
Asian (%)	7.0
Hispanic (%)	11.9
Mean Distance Moved	600.9
N	1,635,000

The table reports summary statistics of a sample of US workers, each of whom is age 25-54, earn at least $15,000 annually, have migrated commuting zones between 2014 and 2019, and work at firms with one establishment in their state. Workers must work in one of 30 states (shown in Figure 1.1) in 2014 but can live in any state in 2019. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.2: Summary statistics: 30-state sample

Variable	Value
Mean Age	35.34
Median Age	34
Male (%)	62.5
Native-born (%)	86.7
White (%)	83.3
Black (%)	8.2
Asian (%)	5.0
Hispanic (%)	13.1
Mean Distance Moved	336.9
N	462,000

The table reports summary statistics of a sample of US workers, each of whom is age 25-54, earn at least $15,000 annually, have migrated commuting zones between 2014 and 2019, and work at firms with one establishment in their state. Workers must work in one of 30 states (shown in Figure 1.1) in both 2014 and 2019. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.3: Summary statistics: Employer Change sample

Variable	**Value**
Mean Age	34.8
Median Age	34
Male (%)	61.8
Native-born (%)	87.6
White (%)	82.8
Black (%)	9.1
Asian (%)	4.4
Hispanic (%)	14.6
Mean Distance Moved	331.4
N	246,000

The table reports summary statistics of a sample of US workers, each of whom is age 25-54, earn at least $15,000 annually, have migrated commuting zones between 2014 and 2019, and work at firms with one establishment in their state. Workers must work in one of 30 states (shown in Figure 1.1) in both 2014 and 2019, and have changed their employer. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.4: Main Sample: Annual Earnings by Percentile

Percentile	Earnings ($)
5	15 040
10	16 760
25	22 730
33	26 420
50	35 820
67	48 840
75	58 020
90	92 070
95	125 100

The table reports the wage distribution of the main sample described in table 1.1. Numerical values are rounded according to Census Bureau disclosure rules.

Table 1.5: Effect of *coworker* on *comove*: Main Sample

$coworker_{ij}$	0.016^{***}
SE	(0.0015)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes
$E[comove\|coworker = 0]$	0.043
Effect size	36.8%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.6: Effect of *coworker* on *comove*: 30 State, Employer Change, and Different Destination Firm Subsamples

Sample	(1) 30 states	(2) Emp. changers	(3) Diff. dest. firms
$coworker_{ij}$	0.040^{***}	0.031^{***}	0.026^{***}
SE	(0.0018)	(0.0016)	(0.0016)
N (pairs)	25,020,000	8,705,000	8,678,000
N (individuals)	462,000	246,000	246,000
Industry-CZ FE	Yes	Yes	Yes
$E[comove\|coworker = 0]$	0.170	0.154	0.154
Effect size	23.4%	19.9%	16.8%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of three regressions where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included in all specifications. Column (1) is a sample restricted to workers who migrate to one of 30 states (see text for details). Relative to column (1), column (2) drops workers who do not change employers. Relative to column (2), column (3) drops pairs of workers who join the same destination firm. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.7: Effect of *coworker* on *comove*: Distance subsamples

Sample	Moved $\geq$ 50 miles	$\geq$ 100	$\geq$ 250
$coworker_{ij}$	.016***	.017***	.019***
SE	(.0018)	(.0021)	(.0031)
N (pairs)	257,000,000	213,200,000	137,000,000
N (individuals)	1,333,000	1,156,000	872,000
Industry-CZ FE	Yes	Yes	Yes
$E[comove\|coworker = 0]$	.033	.032	.030
Effect size	47.0%	52.9%	66.2%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Each column includes only those workers who have migrated at least the specified distance. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included in all specifications. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.8: Effect of *coworker* shared tenure on *comove*

coworker	0.0103***
	(0.0011)
coworker_qtrs	0.00083**
	(0.0003)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes
Mean *coworker_qtrs*	6.9
$E[comove\|coworker = 0]$	0.043
Effect size (*coworker*)	23.8%
Effect size (*coworker_qtrs*)	1.9%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable *coworker_qtrs* is a variable containing the number of quarters the pair of workers have worked together if they are coworkers. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.9: Effect of *coworker* on *comove*: Heterogeneous specification, race and ethnicity

coworker	0.014***
	(0.002)
coworker * both white	0.0029***
	(0.001)
coworker * both black	0.0012
	(0.003)
coworker * both asian	−0.0042*
	(0.002)
coworker * both hispanic	0.0050**
	(0.002)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.10: Effect of *coworker* on *comove*: Heterogeneous specification, gender

coworker	0.016***
	(0.002)
coworker * both male	−0.000 46
	(0.001)
coworker * both female	−0.0011
	(0.001)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.11: Effect of *coworker* on *comove*: Heterogeneous specification, place of birth

coworker	0.015***
	(0.002)
coworker * both native	0.000 52
	(0.002)
coworker * both immigrant	−0.000 40**
	(0.002)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. "Native" is defined as U.S. place of birth. "Immigrant" is defined as non-U.S. place of birth. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.12: Effect of *coworker* on *comove*: Heterogeneous specification, age

coworker	0.016***
	(0.002)
coworker * both 25-34	−0.0017
	(0.001)
coworker * both 35-44	0.0027***
	(0.002)
coworker * both 45-54	0.0091**
	(0.004)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.13: Effect of *coworker* on *comove*: Heterogeneous specification, pre-migration wage

coworker	0.014***
	(0.001)
coworker * both wage $\leq$ 50th pctile	−0.000 76
	(0.001)
coworker * both wage 50-90th pctile	0.0013*
	(0.001)
coworker * both wage $>$ 90th pctile	0.035**
	(0.01)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Wage percentiles are reported in Table 1.4. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.14: Effect of *coworker* on *comove*: Excluding firms that shrink or grow by more than 30%

Sample	All Industries
	$comove_{ij}$
$coworker_{ij}$	0.0133***
SE	(0.0006)
N (pairs)	129,900,000
N (individuals)	944,000
Industry-CZ FE	Yes
$E[comove\|coworker = 0]$	0.047
Effect size	28.5%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Workers who work in firms whose employment decreases or increases by more than 30% between 2014 and 2019 are excluded. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table 1.15: Effect of *coworker* on *comove*: Movers and non-movers

Sample	Movers and Non-movers
	$comove_{ij}$
$coworker_{ij}$	0.00060***
SE	(0.00005)
N (pairs)	457,100,000
N (individuals)	2,145,000
Industry-CZ FE	Yes
$E[comove\|coworker = 0]$	0.00083
Effect size	72.0%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, and work at firms with one establishment in their state in 2014. Workers are included regardless of migration status. The regression is estimated on a 10% random sample of individuals meeting the sample restrictions, prior to constructing dyadic pairs, due to computational burden. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

CHAPTER 2

Labor Specialization and Search Frictions

> The owner of an isolated factory, even if he has access to a plentiful supply of general labour, is often put to great shifts for want of some special skilled labour; and a skilled workman, when thrown out of employment in it, has no easy refuge.
>
> Alfred Marshall, 1890

2.1 Introduction

Since David Ricardo, economists have emphasized that specialization according to comparative advantage and subsequent trade raises productivity and aggregate output. In the modern global economy, the deepening of occupational and task specialization—facilitated by technological progress, multinational supply chains, and increasingly granular skill requirements—has been a key driver of long-run growth. Yet from a macroeconomic perspective, this trend may come with a cost: in labor markets, moreso than in goods markets, reallocating productive factors requires costly and time-consuming search. If greater specialization raises the difficulty of matching workers to jobs, it could reduce aggregate matching efficiency, contribute to higher frictional unemployment, and slow labor reallocation after shocks. Such frictions may manifest as persistent shifts in the Beveridge curve, weaker recoveries from recessions, or a more sluggish response of employment to structural change.

This raises the central question of this paper: Does rising specialization in tasks demanded by firms and skills held by workers reduce the efficiency of labor market matching, and if so, by how much? In the language of the Diamond-Mortensen-Pissarides labor search framework, can the composition of the labor force and vacancies alter the aggregate matching efficiency

parameter that many macroeconomic models take as given? Understanding this relationship is essential for both business cycle analysis and long-run growth theory: it speaks to whether the productivity benefits of specialization are partly offset by slower adjustment to shocks, and whether there is a wedge between the specialization level that is privately optimal for firms and that which maximizes social welfare.

Indeed, there may be theoretical reasons to suspect that increasing job specialization leads to greater search frictions in labor search and matching models, implying reduced worker flows into or out of employment. Firms with very specific skill requirements may find it more difficult to fill vacancies—workers with those skills may be rare, or already in a well-matched job. Likewise, displaced workers with highly specialized skills may have difficulty finding new jobs that match those skills well. At the same time, employers and employees in highly specialized matches may find it more costly to separate, given the lack of close substitutes.

In this paper, using online vacancy data from Burning Glass Technologies, I construct measures of labor specialization based on the rarity of skills associated with a job. Standard U.S. vacancy data, such as the Job Openings and Labor Turnover Survey (JOLTS), do not provide job-level skill requirements. By contrast, Burning Glass data offer rich information on the skills demanded by firms, allowing specialization to be measured directly. I combine these data with hiring flows and unemployment from the Current Population Survey (CPS) to estimate labor matching efficiency, an object at the center of standard macroeconomic labor search models in the Diamond-Mortensen-Pissarides framework, at various occupational disaggregation levels. I then examine the relationship between occupational matching efficiency and labor specialization, finding evidence consistent with the notion that more specialized occupations have decreased matching efficiency (or greater matching frictions).

Beyond its implications for individual firms and occupations, this relationship has important consequences for aggregate labor market performance and the broader macroeconomy. If there is indeed a tradeoff between the productivity gains from specialization and the higher frictional unemployment associated with reduced matching efficiency, an effect that has not been directly documented prior to this paper, then the structure of the labor market itself becomes a determinant of aggregate unemployment and output volatility. In standard macroeconomic search models, matching efficiency is typically treated as exogenous; my results suggest it may be systematically linked to the economy's occupational composition. This opens the door to incorporating endogenous matching efficiency into calibrated macroeconomic models to assess the consequences of shifts in specialization, whether driven by technological change, sectoral shocks, or policy reforms. Quantifying this relationship can inform labor market policies—such as retraining subsidies or mobility incentives—with the aim of preserving the productivity benefits of specialization while mitigating the costs

of slower reallocation.

The structure of the paper is as follows. In section 2, I discuss the paper in relation to relevant literature. In section 3, I detail the Burning Glass online vacancy data and the construction of a measure of job generality. In section 4, I combine vacancy data with unemployment and hires data from the CPS, estimate matching functions at various occupational classification levels, and find the correlation between matching efficiency and job generality. Section 5 concludes and discusses future directions.

2.2 Literature

To the author's knowledge, this paper is the first to estimate matching functions with Burning Glass or similar online vacancy data. Additionally, this paper leverages the richness of Burning Glass vacancy data to construct new measures of specialization, which previous vacancy papers have been unable to do, and links them to matching frictions, a connection previously unexplored.

Past literature that has estimated matching functions have typically used vacancy data from JOLTS (since 2000)—see Petrongolo and Pissarides (2001) for a survey. Prior to that, they typically used indices constructed from newspaper job postings that were known to be noisy. There is also a precursor to the current online vacancy data called HWOL, by the Conference Board (which constructed the original newspaper indices), that began to compile online newspapers and job boards, but it has incomplete coverage compared to Burning Glass and does not have any skills data. The rich data used in this paper offer not only vacancy counts, but also a granular look at the skills demanded by firms, information that was previously unavailable.

An older sectoral shifts literature (Lilien (1982), Abraham and Katz (1986)) posited that if labor demand changes differentially across industries, a constant level of aggregate demand can still lead to unemployment because it takes time to reallocate across jobs in wholly different sectors. A modern iteration of this work is a literature on mismatch in labor markets (Shimer (2007), Sahin et al. (2014)), which attempts to explain inward shifts of the Beveridge curve corresponding to declines in matching efficiency. However, this work is based on vacancy data that includes only counts and has no skills data, which is included in this paper.

There is also recent work on the geography of job tasks (Atalay et al. (2021)) that studies how job tasks and specialization vary across geography and city size. This builds on an older literature positing that the division of labor is limited primarily by coordination costs (Becker and Murphy (1992)), rather than the Smithian notion of the extent of the market (Smith

(2000)). While related, these literatures focus more on labor turnover across geography and does not attempt to engage with the labor market matching process and frictional unemployment as this paper does. Similar to Atalay et al. (2021), this paper constructs a measure of job specialization made possible by the advent of granular job-level task demand, improving on older work (Tian (2018)) that used the number of distinct job titles as a measure of labor specialization.

Lastly, online job vacancy data has become relatively recently more accessible, and a number of recent papers have used such data, whether from Burning Glass or other similar firms, to study a range of topics including job skill requirements, employer concentration, wage rigidity, etc.: Hershbein and Kahn (2018), Deming and Kahn (2018), Azar et al. (2020), Hazell and Taska (2019), Marinescu and Wolthoff (2020).

2.3 Labor Specialization and Generality

2.3.1 Data

Burning Glass Technologies is a software analytics firm that collects and analyzes labor market vacancy postings, scraping them from online job boards such as Indeed or Monster, and then cleaning and deduplicating the postings. The data include the near universe of online job vacancies in the United States, from 2007 to 2020. In this paper, I will use data from 2010 to 2019[1].

The data primarily consist of job postings at the establishment level. Variables include job title, job posting date, firm name, industry/occupation at the SOC 6-digit level, firm location at the zip code level, job description and skill requirements, education level preferred/required, experience level preferred/required, posted wages/salaries, etc. The availability of these variables vary across postings—almost all postings will have a job title, but relatively few will have a wage posted, for instance. While the data does appear to overrepresent certain industries, in particular those that require greater skill, and underrepresent others, the representativeness of the data does not appear to have changed over time and track other sources reasonably closely (Hershbein and Kahn (2018)).

Unfortunately, variables that the data do not have include job posting time and duration. Instead, each job posting is associated with a single date.

My full ten-year sample consists of about 230 million job observations, excluding jobs that are coded as being internships. For purposes of computational speed, all results in

[1]There is only partial data from years before 2010. There is also complete data from 2020, which I omit in this draft because of the unusual labor market activity from the Covid-19 pandemic.

this draft of the paper are computed using a 1% random sample of roughly 2.2 million job observations and 20 million job-skill observations, which corresponds to an average of about 9.3 skills listed per job.

2.3.2 Measurement

In this section, I detail the construction of a measure of generality (the inverse of specialization) at the skill, job, and occupation levels. I also consider some alternative measures of generality in section 2.3.2.4.

2.3.2.1 Skill Generality

The skills data consist of about 20 million job-skill observations, of which there are roughly 14,000 distinct skills. I drop skills that appear in fewer than ten job vacancies. For each of the remaining skills, I construct a frequency measure equivalent to the percentage of job listings in which that skill appears. I interpret the most frequent skills as the most general, with the least frequent skills being the most specialized. Thus this metric is better termed a measure of skill *generality* rather than *specialization*, because a higher value indicates more general skills and a lower value more specialized skills.

The 25 most general skills by this measure is displayed in Table 2.1, with the corresponding percentage of jobs requiring that skill. For instance, the most common skill listed is "communication skills," appearing in about 3% of all job postings, followed by "customer service" appearing in about 1.7% of job listings. These can be skills that can be interpreted as broader categories (i.e. "sales" or "research"), but specific software skills such as "Microsoft Excel" also appear.

A cumulative distribution of the number of skills is given in Figure 2.1, showing that about 1000 skills (∼7%) account for roughly 80% of the total skill postings. Thus, while there are many distinct skills required across all jobs, the vast majority of jobs require comparatively common ones.

2.3.2.2 Job Generality

Next, I construct a measure of generality at the job level. Each job is associated with some number of skills. I then take the average of the frequency of skills associated with each job. For instance, if a job has two skills required, and the frequencies of those skills among job listings in the dataset are 2% and 6%, then the average frequency of skills for this job is 4%. Again, a higher value indicates less specialization (the skills required for this job are

overall quite common), and a lower value more specialization (the job demands skills that few other jobs do).

Figure 2.2 shows the distribution of values of job generality and the density of jobs associated with that value. The long right tail indicates that a few jobs have very general skills, but the vast majority of jobs have average frequencies of skills somewhere between 0 and 1%.

2.3.2.3 Occupation Generality

Finally, I aggregate the jobs into occupations, taking the average generality of each job in that occupation. Jobs are individual, establishment-level labor contracts (ex: a particular chaired professorship at the University of Michigan), while occupations are broader categories of jobs (ex: all economics professors). I do those at the SOC 2-digit, 3-digit, 5-digit, and 6-digit levels. At all levels, I restrict the analysis to occupations which have at least ten job postings associated with them. Not doing so results in obscure occupations with very few jobs (ex: wood patternmakers) showing up at fine levels of occupational disaggregation in the results below.

The 23 SOC 2-digit occupations are ranked in order from most general to least general in Table 2.2. These make some intuitive sense: the three most general occupations are sales, office and administrative support, and food preparation, which we might consider occupations that generally require less training. The most specialized occupations are construction, healthcare, and computer/mathematical occupations, which we think of as requiring many years of education or training.

The top 10 most general jobs and least general jobs at the SOC-6 digit level are reported in tables 2.4 and 2.6, respectively. The least general (most specialized) SOC 6-digit occupations include many medical occupations, such as anesthesiologists, orthodontists, and surgeons. This is unsurprising, given the high degree of training that such jobs require. The most general SOC 6-digit occupations include many common, relatively low-skilled occupations, such as food preparation workers, restaurant and coffeeshop waitstaff, customer service representatives, and bank tellers.

2.3.2.4 Number of Skills and an Alternative Measure of Generality

While I use the frequency of skills associated with job vacancies to compute an index, the number of skills listed for a job can vary widely. How to interpret the number of skills in terms of the specialization level of a job is not necessarily clear. A job with only one skill could be either very general or very specialized depending on whether that skill was

something common like "communications skills," or something obscure like "plastic surgery." Thus number of skills could be an imprecise metric for specialization.

Figure 2.4 shows a histogram of the number of skills associated with various jobs. The vast majority of jobs have under 20 skills, with the mean number of skills per job being about 9.3. The jobs that have more than 20 skills tend to be informational technology jobs, with requirements for various specific software skills largely responsible for driving up the total skill count.

A scatterplot of number of skills vs. the job generality measure constructed above is provided in Figure 2.5. There is a clear negative correlation between number of skills and job generality, as we might expect: jobs with many skills listed tend to be more specialized.

However, while a job with few skills listed could be interpreted as either specialized or general, a large number of skills listed is likely correlated with more specialization. While taking the average of skill frequency as a metric for job generality mitigates some of the concern by normalizing the number of skills, it should be noted that the more skills an employer lists, the more likely those additional skills are to be obscure ones–the generality of the marginal skill is likely to be decreasing. To some extent, however, we also *do* want to capture the notion that an employer that demands more distinct types of skills is likely demanding more specialized labor. But we want to avoid the concern that a firm listing more skills will mechanically decrease the generality of the job.

Thus, in order to account for the number of skills listed, I construct an alternative measure in which I fix the number of skills, taking the average frequency of only the first 15 skills listed per job. If a job has fewer than 15 skills listed, then this measure simply takes the average of the available skills.

Compared to the original measure, this alternative measure will mechanically raise the generality level of jobs with more than 15 skills, since the least frequent skills are being eliminated from the index.

Figure 2.3 shows the distribution of values of the alternative measure of job generality and the density of jobs associated with that value. While not substantially different in character from Figure 2.2, a histogram of the original job generality measure, Figure 2.3 shows a clear rightward shift in the distribution, i.e. a shift of density towards more jobs with higher generality, as expected.

Aggregating the alternative measure to the occupational level generally raises the generality level across all occupations, though the increase is larger for some occupations than others. For example, the generality of computer and information technology occupations sees a greater increase than other occupations, because the jobs that tend to specify many skills tend to be in these occupations.

Table 2.3 shows all occupations ranked from most to least general using the new measure of generality. There are only a few differences compared with Table 2.2. Under the new measure, legal occupations, life, physical, and social science occupations, and computer and mathematical occupations have gone up in generality slightly more than other occupations, increasing their rank by one or two places. This is indicative of these occupations having jobs more likely to require a high number of skills.

Tables 2.5 and 2.7 show the 25 most and least general occupations at the SOC-6 level under the alternative measure, corresponding to Tables 2.4 and 2.6. While there are some shifts among the jobs in both sets of rankings under the new measure relative to the original measure, the occupations listed are mostly the same.

A third possible alternative measure is fixing the number of skills at the occupational level directly, i.e. only taking the 20 most frequent skills in each occupation. However, for a couple of reasons, I prefer the above alternative measure, where I account for the number of skills at the job level before aggregating to the occupational level. First, occupations can consist of many different types of jobs, and particularly at the broader occupation classifications, can subsume many distinct types of job skills. For example, management occupations consist of jobs that between them have hundreds of distinct skill requirements. Thus, accounting for the number of skills at the job level first provides greater variation in the measure. Second, one of the primary advantages of this particular dataset is the richness of skill requirements at the job level. Aggregating at the job level first accounts for the number of skills but without eliminating as much information as accounting for the number of skills directly at the occupational level.

2.4 Matching Efficiency and Labor Specialization

2.4.1 Matching Function Modeling and Estimation

At the core of the standard Diamond-Mortensen-Pissarides labor search framework is the matching function, which takes as inputs the stock of vacancies and unemployment and outputs the flow of matches (or hires) in a given time period:

$$mL = m(uL, vL)$$

where L is the size of the labor force, u the unemployment rate, v the vacancy rate, and m the matching rate.

The matching function is typically specified as increasing in both arguments, concave, and homogeneous of degree one (or constant returns to scale). Empirical evidence supports

a Cobb-Douglas specification for the matching function's functional form (see Petrongolo and Pissarides (2001) for a survey). In this paper I impose this restriction, resulting in the following functional form, with matching parameter m and vacancy share $\alpha \in (0, 1)$ for occupation i at time t:

$$h_{it} = m_{it} v_{it}^{\alpha} u_{it}^{1-\alpha} \tag{2.1}$$

The matching parameter m can be interpreted similarly to total factor productivity in a Cobb-Douglas production function. It can be viewed as a residual: variation in the number of matches (hires) not accounted for by variation in vacancies and unemployment. Additionally, matching efficiency and search frictions can be viewed as inverses of each other—a higher matching efficiency implies less search friction and vice versa. The vacancy share parameter α is the elasticity of matches with respect to vacancies.

Furthermore, this Cobb-Douglas functional form suggests the following log-linear regression specification:

$$\ln\left(\frac{h_{it}}{u_{it}}\right) = \ln(m_{it}) + \alpha \ln\left(\frac{v_{it}}{u_{it}}\right) + \epsilon_{it} \tag{2.2}$$

With monthly-frequency data, the matching function is typically specified with lagged vacancy and unemployment variables (i.e. with $\ln(v_{it-1}/u_{it-1})$), with the notion that matching takes longer than a month—see Blanchard and Diamond (1989). However, my data are at the quarterly level, so whether or not to used lagged vacancies and unemployment is somewhat debatable—Haltiwanger and Davis (2013) show using JOLTS that average vacancy duration in the US is between 2 and 4 weeks, while unemployment durations can often be as long as several months. For completeness, I estimate (2) using both specifications, contemporaneous and lagged.

By combining data on matches and unemployment from the Current Population Survey (CPS) with vacancy counts from the Burning Glass data, this specification allows for the estimation of matching efficiency m (the primary parameter of interest in this paper) as well as the vacancy share parameter α. Specifically, a regression of log hires over unemployment on log vacancies over unemployment will produce estimates of the log matching efficiency (the constant term) and the vacancy share (the coefficient on log vacancies over unemployment). I do this at various occupational classification levels. I detail the construction of the relevant data below in Section 2.4.2.

Finally, I study the relationship between the estimated matching efficiency with the measure of occupational generality detailed previously, regressing the former on the latter. A positive correlation between occupation generality and matching efficiency would suggest

that more specialized jobs have decreased matching efficiency, while a negative correlation would suggest that more specialized jobs have higher matching efficiency. The results, which I present below in Section 2.4.3.2, suggest a positive correlation.

2.4.2 Unemployment and New Hires from CPS Data

Using CPS data from the same 2010-2019 time period, I construct measures of the stock of unemployment and the flow of hires at the SOC 2, 3, 5, and 6-digit occupational levels. To construct the stock of unemployed workers, I aggregate the number of individuals who are coded as either experienced unemployed workers or as unemployed new workers, weighted by their person-level CPS sample weights.

To construct the flow of new hires by occupation, I primarily consider three types of workers, following Hershbein (2017). The CPS is a monthly survey, so I compare worker responses across months. First, workers who are working in the current survey month but were unemployed in the previous survey month are considered new hires for the current month.

Second, workers who are working in the current survey month but were reported as being out of the labor force in the previous month are also considered new hires. Note that this hiring from out of the labor force (rather than from unemployment) is a well-known issue (see Elsby et al. (2015) for a brief conceptual discussion). Thus we might think of these workers as discouraged workers or workers marginally attached to the labor force. Note that the stocks of unemployed workers constructed above do not include this group of workers that are finding jobs from out of the labor force.

Finally, workers who are working in two consecutive months but report a different employer in the second month are considered new hires in the second month. These are essentially job-to-job transitions, workers who either skip the unemployment state entirely or were unemployed for short enough periods to not be captured in the monthly survey.

A potential category of new hire that these three categories may leave out is a worker who stays at the same employer but is in a different job (perhaps a promotion or internal hire). However, the CPS data do not reliably allow for distinguishing such cases, and one would have to take a conceptual stand on whether such cases constitute new jobs or hires.

Particularly at the finer occupational levels, there can be occupations that see few or no hires or unemployed workers in any particular month. Thus to obtain larger sample sizes, I then aggregate the vacancies, unemployment and new hire measures to the quarterly level, similar to Sahin et al. (2014).

2.4.3 Results

I find evidence consistent with the notion that more specialized occupations have lower matching efficiency than more general occupations. I first discuss results from the matching efficiency estimation, then results from regressing matching efficiency on occupational generality.

2.4.3.1 Matching Efficiency Estimation

The results from the matching efficiency estimation are roughly in line with previous estimates in the literature. Table 2.8 shows the results from the matching estimation using contemporaneous vacancies and unemployment for the 22 civilian occupations at the SOC 2-digit level[2]. Table 2.9 shows the corresponding estimates using lagged vacancies and unemployment.

Due to the number of occupations at the SOC 3, 5, and 6-digit levels numbering in the hundreds, I leave out the detailed results, although scatter plots containing the matching efficiency estimates at these levels are presented in Figures B.5 through B.13. The contemporaneous estimates of α, the vacancy share parameter or the elasticity of matches with respect to vacancies, range from 0.234 (for Protective Service Occupations) to 0.607 (Management Occupations). These are roughly similar to previous findings in the literature at the aggregate level, many of which fall in a range from .3 to .6 (see Petrongolo and Pissarides (2001) for a survey).

The contemporaneous estimated matching efficiencies at the SOC-2 level range from 0.297 to 1.940, with the majority close to 1. The three most match-efficient occupations are educational, education, training, and library occupations, personal care and service occupations, and farming, fishing, and forestry occupations. The three least match-efficient occupations are computer and mathematical occupations, architecture and engineering occupations, and healthcare practitioner and technical occupations.

The estimates using lagged vacancies and unemployment in Table 2.9 are similar to the contemporaneous ones. Some occupations now have slightly higher match efficiencies (computer and mathematical occupations), while others have slightly lower match efficiencies (educational occupations, farming occupations), but overall the results are not substantially different.

[2] "Military occupations" is left out because workers in the armed forces are treated separately by the BLS.

2.4.3.2 Matching Efficiency and Occupational Generality

At the SOC-2 level, Figure B.1 graphs the estimated matching efficiencies against the main measure of job generality constructed above, and fits a least-squares line to the scatterplot. The plot shows a mild positive correlation, suggesting that more general jobs have a higher matching efficiency.

Additionally, graphs at the SOC-2 level using the alternative measure of generality and using the lagged matching function specification are presented in Figures B.3 (alternative, contemporaneous), B.2 (main, lagged), and B.4 (alternative, lagged).

Similar plots for estimates at the SOC 3, 5, and 6-digit levels are shown in Figures B.5 through B.16. Generally, it appears that the positive relationship between match efficiency and job generality becomes progressively weaker as the occupational classification becomes increasingly fine.

The estimates of matching efficiency at the finer occupational levels can be imprecise, so in the 3, 5, and 6-digit figures I have dropped some observations in which the matching efficiency estimates are either unable to be computed because the number of observations for a particular occupation is too small, and cases in which the matching efficiency estimate is within two standard errors of zero.

Finally, a summary of the regression results at each of the SOC levels are reported in Tables 2.10 and 2.11. Generally, the estimates of matching efficiency (which are inputs into this final regression) are less precise at finer occupational classification levels, the sample size (total number of occupations) for the final regression is larger for finer occupational levels, which explains the smaller standard errors at the 5 and 6-digit SOC levels. So there is a tradeoff between the precision of the matching efficiency estimate and the precision of the final regression.

The estimates that include the alternative measure of generality as well as the matching efficiency from lagged vacancies and unemployment, while generally still indicating a positive correlation between matching efficiency and generality, are somewhat more muted as a whole and less precisely estimated.

Interpreting the main estimates at face value would suggest that at the SOC-2 level, a 1% increase in occupational generality is associated with a .99% increase in matching efficiency. At finer occupational levels, a 1% increase in occupational generality is associated with roughly a third of a percentage point increase in matching efficiency.

2.5 Conclusion

In this paper, I construct a measure of labor specialization using online vacancy data, and study the relationship between matching efficiency and job specialization at various levels of occupational disaggregation. The results in this paper document which occupations tend to be more or less specialized than others, and provide evidence that more specialized occupations have some degree of reduced matching efficiency or increased search frictions. That is, occupations consisting of jobs with more specialized skills tend to have more difficulty matching relative to others.

Several promising avenues build on the core results of this paper. First, a more structured modeling approach offers a productive direction. There have been some attempts to micro-found the modeling of the matching function in the literature (Lagos (2000), Burdett et al. (2001)), but none that have received widespread acceptance or that have added empirical viability. A model that would endogenize matching efficiency arising out of specialization from comparative advantage would formalize the ideas in this paper. Integrating the estimated relationship between specialization and matching efficiency into calibrated macroeconomic frameworks that endogenize matching efficiency would allow quantitative evaluation of how shifts in occupational composition, whether driven by technological change, sectoral shocks, or policy interventions, may affect aggregate unemployment, vacancy durations, and output. Embedding this channel into structural models would also enable welfare analysis of counterfactual policies, including retraining subsidies, mobility incentives, and measures to encourage broader skill acquisition, and could help policymakers navigate the tradeoff between productivity gains from specialization and slower labor reallocation.

Moreover, models combining search frictions with spatial equilibrium, as in Kline and Moretti (2013), would be important in studying agglomeration effects, with important implications in urban economics.

A second avenue is to enhance the empirical matching function estimation. The main regression on log hires, vacancies, and unemployment estimated in this paper is similar to baseline regressions in classic papers such as Blanchard and Diamond (1989). More recently, and with better data (JOLTS did not exist prior to 2000), the literature has utilized some more advanced methods, such as more flexible CES specifications of the matching function, which tend to support the Cobb-Douglas specification, (see Petrongolo and Pissarides (2001)), or imposing an ARMA(p,q) structure matching efficiency and utilizing a GMM estimator, as proposed by Borowczyk-Martins et al. (2013). The latter argues that endogeneity issues can bias OLS estimates—in particular, shocks to unobserved matching efficiency may affect firm demand for labor, changing their vacancy posting behavior.

A third extension involves leveraging the geographic detail in the vacancy data to analyze how specialization and matching efficiency vary across local labor markets. This would also allow me to treat labor markets at the occupation-geography level instead of at the occupation level. Given CPS sample size constraints, this may necessitate a tradeoff between fineness of occupation and geography—the finer the occupation classification, the less geographical variation can be considered while preserving statistical precision. To move in this direction, unemployment data at the county and MSA level from the Local Area Unemployment Statistics (LAUS) provided by the BLS would be useful.

Indeed, descriptive patterns support the importance of geography. Table B.1 shows variation in the number of vacancies and the average job-level generality (both main and alternative measures) across the 50 US states and Washington, DC. Table 2.12 shows variation across MSA types, which are geographic regional divisions defined by the US Office of Management and Budget and used by the Census Bureau. There are two primary MSA types: metropolitan statistical areas and micropolitan statistical areas. The former have population of at least 50,000 and include major cities, while the latter have population between 10,000 and 50,000.

The data show that the larger states tend to have more vacancies, which is to be expected—for example, California, Texas, and New York have the highest numbers of vacancies. Additionally, metropolitan statistical areas have almost 20 times as many vacancies as micropolitan statistical areas. The average number of skills required by jobs also appears to be higher in larger states, and higher in metropolitan statistical areas than micropolitan ones: about 9.5 skills in metropolitan areas vs. about 7 skills per job in micropolitan areas.

However, there is a not a clear relationship between state size and job generality, and the MSA variation shows that metropolitan areas actually have slightly more general jobs on average than micropolitan statistical areas. While one might expect metropolitan areas to have more specialized jobs, there is also evidence that innovative (and presumably more specialized) jobs result in the creation of additional service jobs (likely less specialized)—see Moretti (2012) and Moretti (2010). For example, many software developers moving to Silicon Valley also means that there is increased demand for baristas, laundry workers, etc. whose services the developers will desire. The net effect on specialization in cities is therefore theoretically ambiguous, and merits further investigation. The data appear to show little difference on net when it comes to job generality, but a greater average number of skills in metropolitan areas.

Finally, incorporating posted wage information from the vacancy data would enable estimating productivity gains from specialization and compare them to the costs of reduced matching efficiency. This would provide a fuller picture of the specialization–frictions trade-

off and strengthen the link between micro-level vacancy patterns and macro-level welfare outcomes.

2.6 Tables

Table 2.1: 25 Most Common Skills and their Frequency

Skill	**% of jobs requiring skill**
Communication Skills	3.24
Customer Service	1.79
Teamwork / Collaboration	1.48
Organizational Skills	1.40
Sales	1.36
Microsoft Excel	1.24
Problem Solving	1.23
Scheduling	1.17
Planning	1.14
Detail-Oriented	1.07
Physical Abilities	1.06
Writing	1.05
Computer Literacy	1.05
Microsoft Office	0.99
Research	0.95
Budgeting	0.82
Written Communication	0.79
Building Effective Relationships	0.79
Project Management	0.73
Customer Contact	0.69
Retail Industry Knowledge	0.68
English	0.67
Multi-Tasking	0.66
Repair	0.65
Creativity	0.63

Table 2.2: All Occupations Ranked from Most to Least General, SOC-2 Level: Main Generality Measure

SOC-2 Code	**Occupation**
41	Sales and Related Occupations
43	Office and Administrative Support Occupations
35	Food Preparation and Serving Related Occupations
33	Protective Service Occupations
27	Arts, Design, Entertainment, Sports, and Media Occupations
55	Military Specific Occupations
11	Management Occupations
13	Business and Financial Operations Occupations
25	Education, Training, and Library Occupations
23	Legal Occupations
37	Building and Grounds Cleaning and Maintenance Occupations
49	Installation, Maintenance, and Repair Occupations
39	Personal Care and Service Occupations
21	Community and Social Service Occupations
53	Transportation and Material Moving Occupations
45	Farming, Fishing, and Forestry Occupations
51	Production Occupations
19	Life, Physical, and Social Science Occupations
17	Architecture and Engineering Occupations
31	Healthcare Support Occupations
15	Computer and Mathematical Occupations
29	Healthcare Practitioners and Technical Occupations
47	Construction and Extraction Occupations

Table 2.3: All Occupations Ranked from Most to Least General, SOC-2 Level: Alternative Generality Measure

SOC-2 Code	**Occupation**
41	Sales and Related Occupations
43	Office and Administrative Support Occupations
35	Food Preparation and Serving Related Occupations
33	Protective Service Occupations
27	Arts, Design, Entertainment, Sports, and Media Occupations
11	Management Occupations
13	Business and Financial Operations Occupations
55	Military Specific Occupations
23	Legal Occupations
25	Education, Training, and Library Occupations
49	Installation, Maintenance, and Repair Occupations
37	Building and Grounds Cleaning and Maintenance Occupations
39	Personal Care and Service Occupations
21	Community and Social Service Occupations
53	Transportation and Material Moving Occupations
19	Life, Physical, and Social Science Occupations
51	Production Occupations
45	Farming, Fishing, and Forestry Occupations
17	Architecture and Engineering Occupations
15	Computer and Mathematical Occupations
31	Healthcare Support Occupations
29	Healthcare Practitioners and Technical Occupations
47	Construction and Extraction Occupations

Table 2.4: Top 25 Most General Occupations, SOC-6 Level: Main Generality Measure

SOC6 Code	**Title**
11-9061	Funeral Service Managers
39-4021	Funeral Attendants
35-9099	Food Preparation and Serving Related Workers, All Other
41-2022	Parts Salespersons
39-1021	First-Line Supervisors of Personal Service Workers
35-9031	Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop
39-2011	Animal Trainers
51-6041	Shoe and Leather Workers and Repairers
43-4051	Customer Service Representatives
37-2021	Pest Control Workers
41-2021	Counter and Rental Clerks
29-1011	Chiropractors
35-3022	Counter Attendants, Cafeteria, Food Concession, and Coffee Shop
43-3071	Tellers
39-9099	Personal Care and Service Workers, All Other
53-6021	Parking Lot Attendants
39-3011	Gaming Dealers
49-2096	Electronic Equipment Installers and Repairers, Motor Vehicles
43-5032	Dispatchers, Except Police, Fire, and Ambulance
13-1074	Farm Labor Contractors
39-4031	Morticians, Undertakers, and Funeral Directors
43-4181	Reservation and Transportation Ticket Agents and Travel Clerks
41-2031	Retail Salespersons
51-9071	Jewelers and Precious Stone and Metal Workers
43-4141	New Accounts Clerks

Table 2.5: Top 25 Most General Occupations, SOC-6 Level: Alternative Generality Measure

SOC-6 Code	**Occupation**
11-9061	Funeral Service Managers
39-4021	Funeral Attendants
35-9099	Food Preparation and Serving Related Workers, All Other
39-1021	First-Line Supervisors of Personal Service Workers
41-2022	Parts Salespersons
35-9031	Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop
39-2011	Animal Trainers
51-6041	Shoe and Leather Workers and Repairers
43-4051	Customer Service Representatives
37-2021	Pest Control Workers
41-2021	Counter and Rental Clerks
43-3071	Tellers
35-3022	Counter Attendants, Cafeteria, Food Concession, and Coffee Shop
29-1011	Chiropractors
39-4031	Morticians, Undertakers, and Funeral Directors
39-9099	Personal Care and Service Workers, All Other
53-6021	Parking Lot Attendants
39-3011	Gaming Dealers
43-5032	Dispatchers, Except Police, Fire, and Ambulance
49-2096	Electronic Equipment Installers and Repairers, Motor Vehicles
43-4141	New Accounts Clerks
13-1074	Farm Labor Contractors
43-4181	Reservation and Transportation Ticket Agents and Travel Clerks
41-2031	Retail Salespersons
41-9099	Sales and Related Workers, All Other

Table 2.6: Top 25 Least General Occupations, SOC-6 Level: Main Generality Measure

SOC6 Code	**Title**
47-2131	Insulation Workers, Floor, Ceiling, and Wall
29-1061	Anesthesiologists
47-3012	Helpers–Carpenters
29-1024	Prosthodontists
29-1066	Psychiatrists
29-1063	Internists, General
29-1064	Obstetricians and Gynecologists
29-1023	Orthodontists
29-1062	Family and General Practitioners
47-2051	Cement Masons and Concrete Finishers
47-2081	Drywall and Ceiling Tile Installers
29-1029	Dentists, All Other Specialists
47-2132	Insulation Workers, Mechanical
29-1065	Pediatricians, General
29-1067	Surgeons
29-1069	Physicians and Surgeons, All Other
29-1022	Oral and Maxillofacial Surgeons
47-3013	Helpers–Electricians
29-1151	Nurse Anesthetists
51-4011	Computer-Controlled Machine Tool Operators, Metal and Plastic
29-1081	Podiatrists
39-9011	Childcare Workers
51-4041	Machinists
51-4121	Welders, Cutters, Solderers, and Brazers
53-3022	Bus Drivers, School or Special Client

Table 2.7: Top 25 Least General Occupations, SOC-6 Level: Alternative Generality Measure

SOC-6 Code	**Occupation**
47-2131	Insulation Workers, Floor, Ceiling, and Wall
29-1061	Anesthesiologists
47-3012	Helpers–Carpenters
29-1024	Prosthodontists
29-1066	Psychiatrists
29-1023	Orthodontists
29-1063	Internists, General
29-1064	Obstetricians and Gynecologists
29-1062	Family and General Practitioners
47-2051	Cement Masons and Concrete Finishers
47-2081	Drywall and Ceiling Tile Installers
29-1065	Pediatricians, General
29-1029	Dentists, All Other Specialists
29-1067	Surgeons
47-2132	Insulation Workers, Mechanical
29-1069	Physicians and Surgeons, All Other
47-3013	Helpers–Electricians
29-1022	Oral and Maxillofacial Surgeons
29-1081	Podiatrists
39-9011	Childcare Workers
29-1151	Nurse Anesthetists
51-4011	Computer-Controlled Machine Tool Operators, Metal and Plastic
51-4121	Welders, Cutters, Solderers, and Brazers
53-3022	Bus Drivers, School or Special Client
51-4041	Machinists

Table 2.8: Matching Function Estimation, SOC-2: Contemporaneous vacancies and unemployment

SOC-2 Occupation	α	α **SE**	**Match Efficiency**	**Match Eff. SE**
Management Occupations	0.607	0.056	1.077	0.037
Business and Financial Operations Occupations	0.486	0.074	0.867	0.043
Computer and Mathematical Occupations	0.363	0.165	0.297	0.168
Architecture and Engineering Occupations	0.582	0.052	0.680	0.032
Life, Physical, and Social Science Occupations	0.566	0.067	1.012	0.045
Community and Social Service Occupations	0.539	0.055	1.453	0.066
Legal Occupations	0.371	0.125	1.219	0.084
Education, Training, and Library Occupations	0.505	0.059	1.940	0.104
Arts, Design, Entertainment, Sports, and Media Occupations	0.475	0.051	1.350	0.063
Healthcare Practitioners and Technical Occupations	0.475	0.086	0.852	0.067
Healthcare Support Occupations	0.405	0.044	1.456	0.092
Protective Service Occupations	0.257	0.065	1.121	0.112
Food Preparation and Serving Related Occupations	0.336	0.037	1.416	0.090
Building and Grounds Cleaning and Maintenance Occupations	0.334	0.041	1.379	0.104
Personal Care and Service Occupations	0.359	0.050	1.663	0.116
Sales and Related Occupations	0.492	0.047	1.259	0.062
Office and Administrative Support Occupations	0.472	0.038	1.274	0.058
Farming, Fishing, and Forestry Occupations	0.234	0.044	1.587	0.204
Construction and Extraction Occupations	0.302	0.029	1.467	0.104
Installation, Maintenance, and Repair Occupations	0.445	0.044	1.055	0.057
Production Occupations	0.401	0.037	1.325	0.089
Transportation and Material Moving Occupations	0.334	0.025	1.277	0.057

Regression of log hires over unemployment on log vacancies over unemployment. The estimate of the constant term is match efficiency, and the estimate of the coefficient on log vacancies over unemployment is α.

Table 2.9: Matching Function Estimation, SOC-2: Lagged vacancies and unemployment

SOC-2	**alpha**	**alpha SE**	**Match Efficiency**	**Match Efficiency SE**
Management Occupations	0.538	0.053	1.091	0.036
Business and Financial Operations Occupations	0.395	0.075	0.867	0.044
Computer and Mathematical Occupations	0.243	0.154	0.444	0.155
Architecture and Engineering Occupations	0.501	0.063	0.710	0.038
Life, Physical, and Social Science Occupations	0.318	0.076	0.947	0.052
Community and Social Service Occupations	0.470	0.062	1.419	0.075
Legal Occupations	0.144	0.140	1.104	0.096
Education, Training, and Library Occupations	0.201	0.091	1.460	0.164
Arts, Design, Entertainment, Sports, and Media Occupations	0.405	0.053	1.311	0.066
Healthcare Practitioners and Technical Occupations	0.432	0.078	0.913	0.060
Healthcare Support Occupations	0.352	0.042	1.384	0.089
Protective Service Occupations	0.220	0.061	1.090	0.107
Food Preparation and Serving Related Occupations	0.292	0.030	1.355	0.073
Building and Grounds Cleaning and Maintenance Occupations	0.286	0.040	1.300	0.103
Personal Care and Service Occupations	0.310	0.049	1.587	0.116
Sales and Related Occupations	0.433	0.043	1.230	0.058
Office and Administrative Support Occupations	0.432	0.037	1.258	0.056
Farming, Fishing, and Forestry Occupations	0.130	0.049	1.141	0.227
Construction and Extraction Occupations	0.256	0.031	1.335	0.110
Installation, Maintenance, and Repair Occupations	0.410	0.042	1.063	0.055
Production Occupations	0.351	0.036	1.250	0.086
Transportation and Material Moving Occupations	0.293	0.021	1.225	0.049

Regression of log hires over unemployment on log vacancies over unemployment, with vacancies and unemployment lagged by one period. The estimate of the constant term is match efficiency, and the estimate of the coefficient on log vacancies over unemployment is α.

Table 2.10: Matching Efficiency and Occupational Generality, Contemporaneous Matching Function Estimation

	Dependent Variable: Log Matching Efficiency (Contemporaneous Vacancies and Unemployment)			
Variable	SOC-2	SOC-3	SOC-5	SOC-6
Occupational Generality (Main Measure)	.99 (.78)	.23 (.32)	.34 (.17)	.31 (.18)
Occupational Generality (Alternative Measure)	.61 (.80)	-.06 (.33)	.26 (.18)	.21 (.20)

Regression of log matching efficiency on occupational generality, with both the main measure of generality and the alternative measure (that uses only the 15 most frequent skills for each job) for various occupational classification levels. Matching efficiency comes from matching function estimation with contemporaneous vacancies and unemployment. Standard errors in parentheses.

Table 2.11: Matching Efficiency and Occupational Generality, Lagged Matching Function Estimation

	Dependent Variable: Log Matching Efficiency (Lagged Vacancies and Unemployment)			
Variable	SOC-2	SOC-3	SOC-5	SOC-6
Occupational Generality (Main Measure)	.83 (.56)	.15 (.30)	.11 (.10)	.18 (.16)
Occupational Generality (Alternative Measure)	.56 (.57)	.04 (.30)	.03 (.10)	.08 (.16)

Regression of log matching efficiency on occupational generality, with both the main measure of generality and the alternative measure (that uses only the 15 most frequent skills for each job) for various occupational classification levels. Matching efficiency comes from matching function estimation with vacancies and unemployment lagged by one period. Standard errors in parentheses.

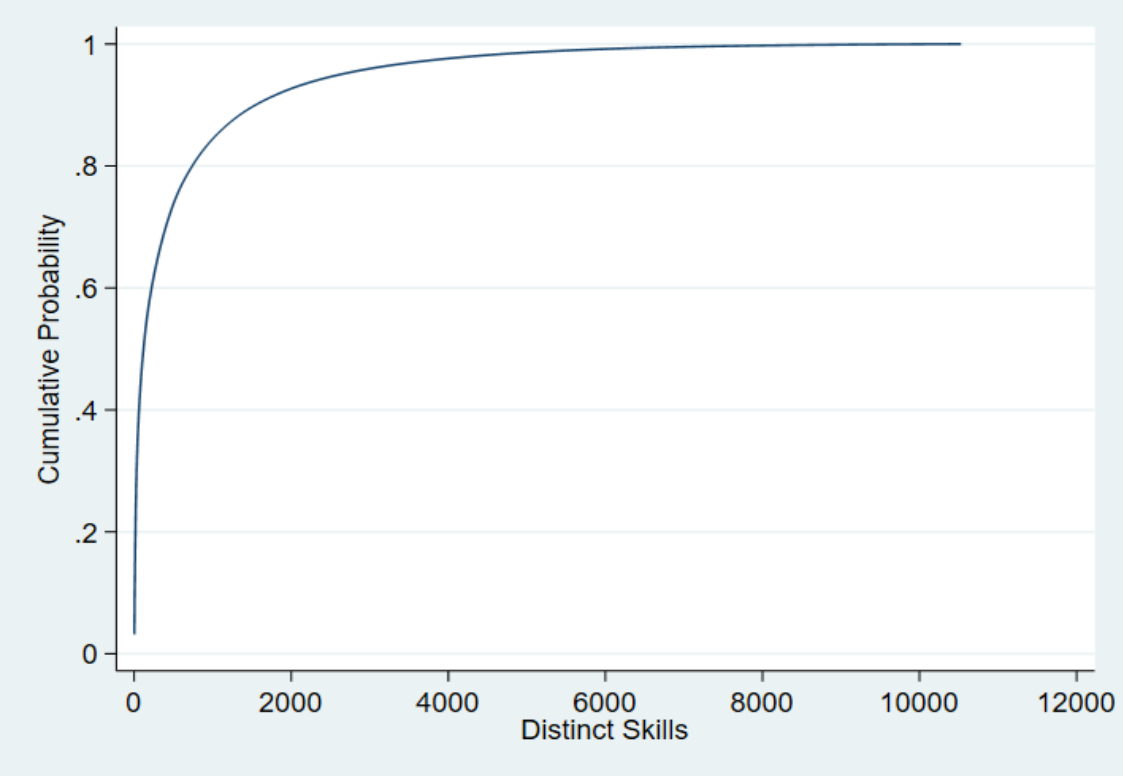
Table 2.12: Vacancy and Generality Variation Across MSA Types, 2010-2019

State	**Total Vacancies**	**Avg Generality (Main)**	**Avg Generality (Alternative)**	**Avg # Skills**
Metropolitan Statistical Area	1881551	.00464	.00492	9.56
Micropolitan Statistical Area	107736	.00460	.00474	7.26

MSAs are defined by the US Office of Management and Budget (OMB) and used by the Census Bureau. Metropolitan Statistical Areas have population greater than 50,000; Micropolitan Statistical Areas have population between 10,000 and 50,000. Generality is measured at the job level, using both the main measure and the alternative measure (which uses only the 15 most frequent skills per job)

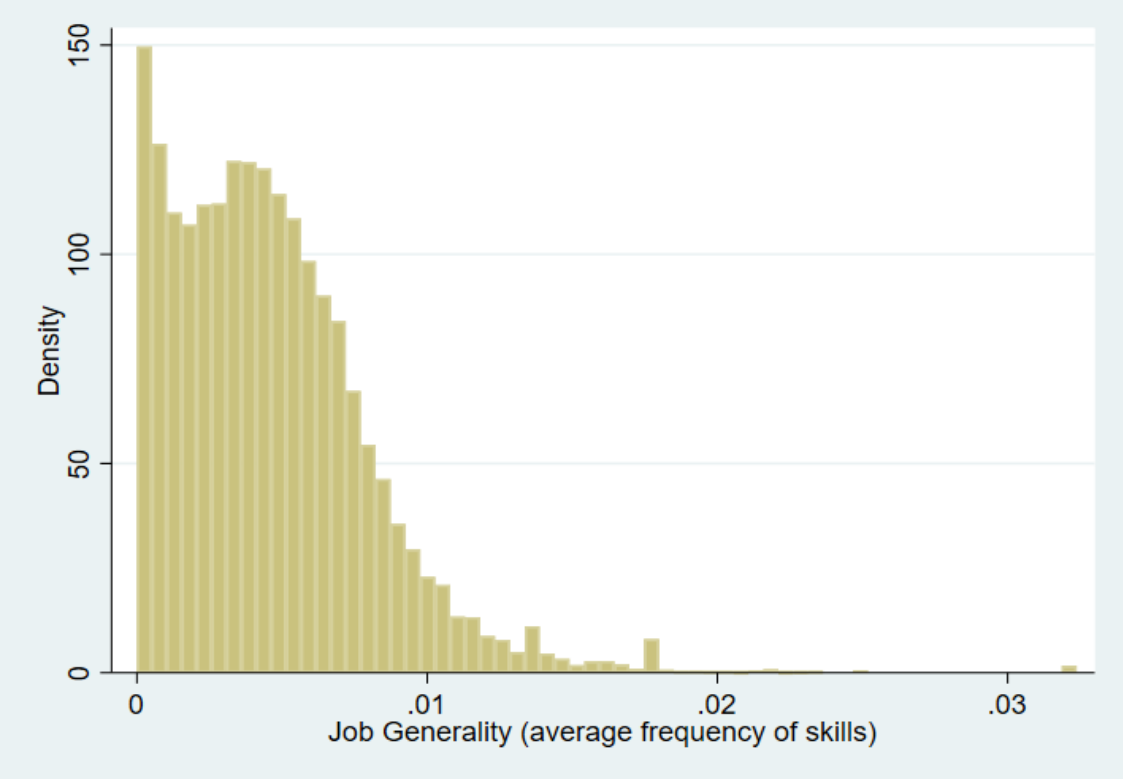
2.7 Figures

Figure 2.1: CDF of Skills



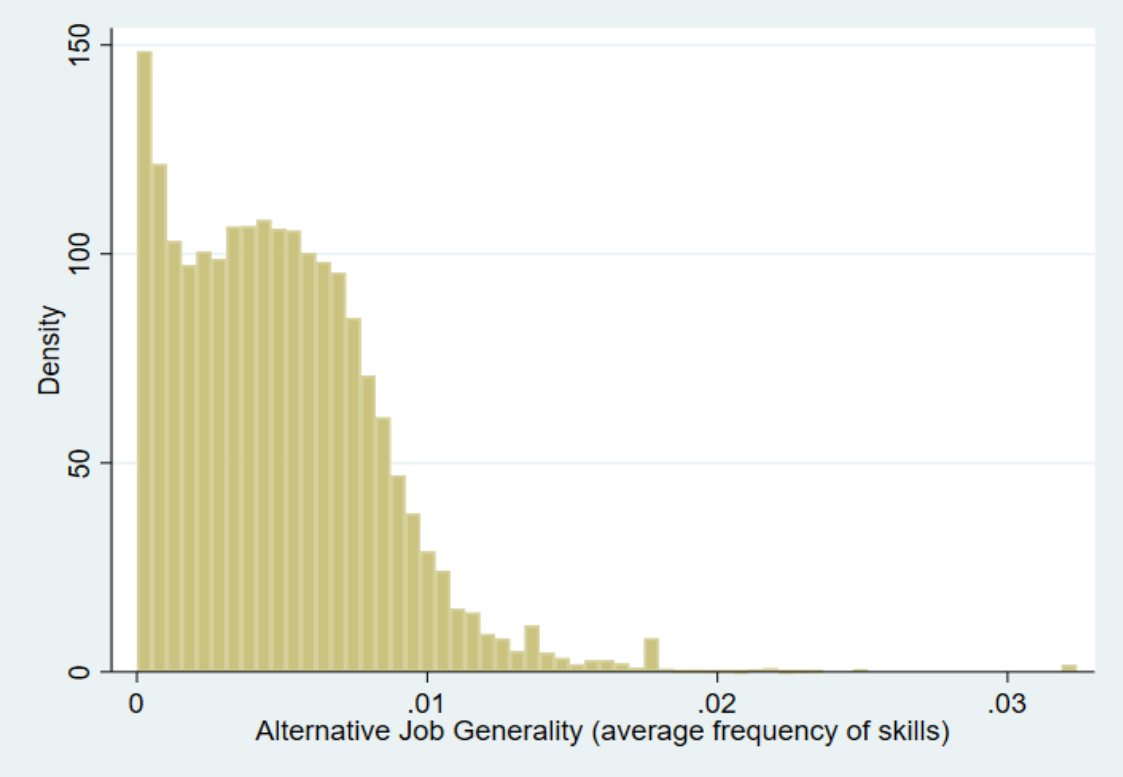
Cumulative distribution of skills. About 1000 distinct skills account for a large proportion ($\sim 80\%$) of job postings. Based on Burning Glass vacancy data.

Figure 2.2: Histogram of Job Generality: Main Generality Measure



This figure displays a histogram of job generality, which is the weighted average of skill frequencies. Based on Burning Glass vacancy data.

Figure 2.3: Histogram of Job Generality: Alternative Generality Measure



This figure displays a histogram of the alternative measure of job generality, which is the weighted average of skill frequencies for only the 15 most frequent skills associated with a job. Based on Burning Glass vacancy data.

Figure 2.4: Histogram of Number of Skills

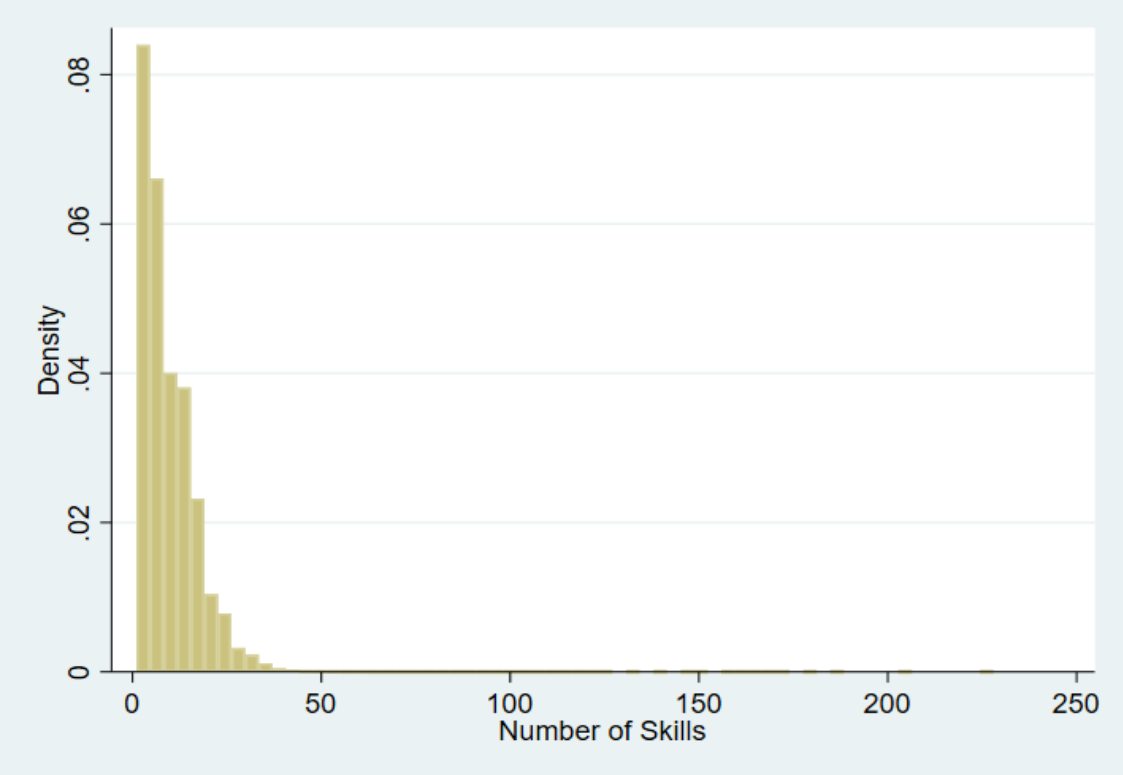
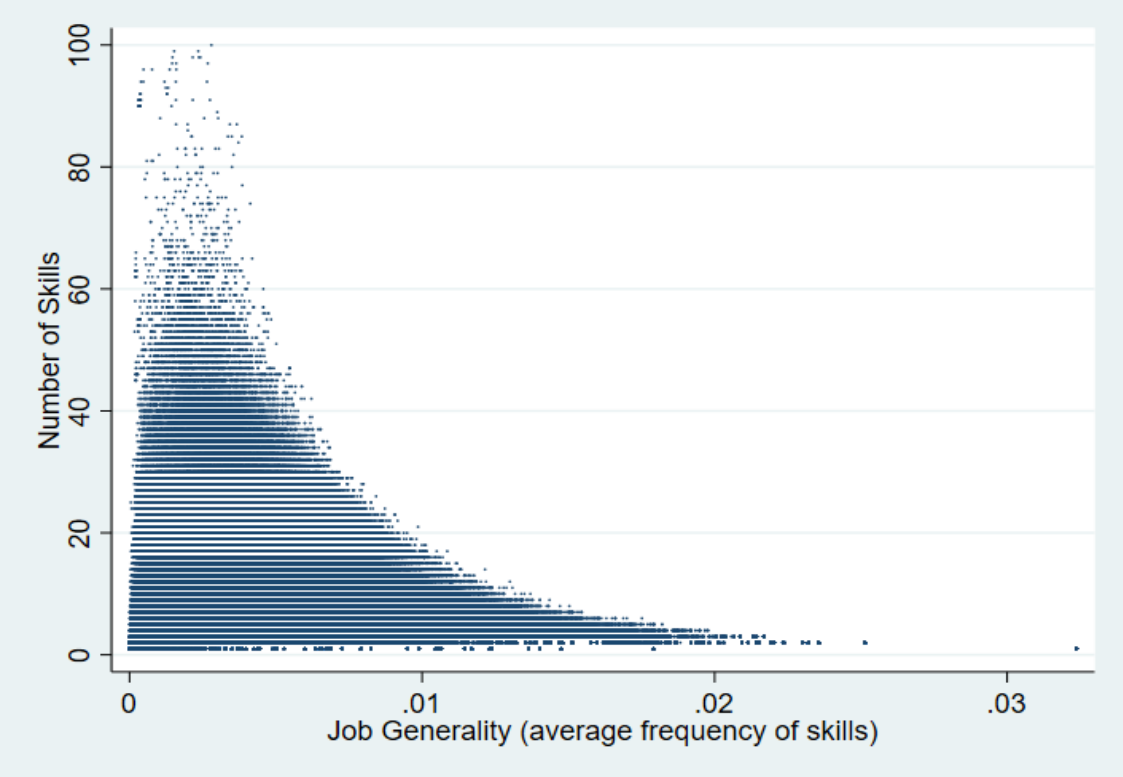


Figure displays a histogram of the number of skills associated with a job, with a mean of 9.3. The vast majority of jobs have fewer than 20 skills. Based on Burning Glass vacancy data.

Figure 2.5: Scatterplot of Skill Generality vs. Number of Skills



Plots the measure of skill generality against number of skills associated with a job. The negative correlation shows that the greater the number of skills required, the less frequent the skills required for that job is.

CHAPTER 3

Residential and Employment Locations in the Post-Pandemic US: Evidence from US Administrative Data

3.1 Introduction and motivation

The COVID-19 pandemic triggered a rapid and widespread shift in work arrangements across the United States. What began as a public health necessity has become a lasting transformation in how and where work is performed. Even as many workplaces have resumed in-person operations, remote and hybrid work models remain a defining feature of the modern, post-pandemic labor market. This shift has significant implications for where people live and where firms source their labor. When work is no longer tethered to a physical location, both workers and employers gain flexibility in making location-based decisions.

For workers, reduced commuting needs potentially expand the set of affordable or desirable residential options, allowing migration to areas with lower housing costs or better amenities. For firms, decoupling employment from geography may broaden access to talent pools, including candidates outside of traditional commuting distances. Together, these dynamics are reshaping the spatial structure of labor markets.

In this paper, I examine how patterns in the geographic relationship between workers' homes and their employers' locations have changed since the start of the COVID-19 pandemic. Using administrative data from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program, I measure the distances between residential and workplace locations over time. Additionally, I assess how these distances vary by industry, gender, and hiring cohort.

I document three main findings. First, I find that mean and median work-home distances have increased since 2020Q2 (which I denote as the start of the pandemic). The increase in distance is largely driven by workers who have been hired since 2020Q2, as opposed to

workers hired before then, suggesting that the geographic labor markets in which employers are searching has expanded as a result of the pandemic. Second, while the distance increase among post-pandemic hires peaked early, there has since been a modest decline, suggesting a re-convergence consistent with return-to-office trends. Third, these patterns are heterogeneous across industries and worker demographics, with the increase in distance is driven primarily by jobs in the information technology industry, and to a lesser extent by jobs in the professional services, finance, administration, and educational services industries. Both men and women saw increases in their work-home distances over time.

These findings suggest that the pandemic durably expanded the geographic footprint of labor markets, particularly along the hiring margin. However, the partial retraction of distances in later years highlights the dynamic nature of these spatial labor market trends. This work adds to our nascent understanding of how technological and organizational shifts—accelerated by a crisis—reshape economic geography.

The rest of the paper proceeds as follows. Section 3.2 describes the literature and discusses how the paper contributes. Section 3.3 describes the LEHD data utilized. Section 3.4 describes the results. Section 3.5 concludes.

3.2 Literature

This paper contributes to a growing body of research on the spatial implications of remote work, particularly in the wake of the COVID-19 pandemic. Early research documented the rise of remote work and its differential adoption across occupations, industries, and demographics (e.g., Barrero et al. 2021; Adams-Prassl et al. 2020; DeFilippis et al. 2020). Subsequent work has focused on productivity outcomes, employee preferences, commuting time savings, and impacts on urban real estate markets (e.g., Aksoy et al. 2023; Gupta et al. 2024; Ramani et al. 2024).

More specifically, this paper relates to recent empirical efforts to document changes in the spatial structure of employment relationships. The closest antecedent is Akan et al. (2024), who use proprietary payroll data to show that home-to-employer distances rose significantly in the post-pandemic period. They also link these changes to broader migration trends, including moves to lower-tax and lower-cost regions, and show that distant employees are more flexibly adjusted on the hiring and separation margins.

This study complements and extends that work by leveraging a different data source—administrative LEHD records—that offers far broader coverage and a longer pre-pandemic time series. The LEHD data capture employment for roughly 95% of the U.S. private sector and provide detailed information on both residential and workplace locations,

enabling precise distance calculations at the individual level. Although the LEHD sample I use is limited in state availability and less timely than proprietary payroll data, it offers a rich and representative view of national labor dynamics, and thus the differences from and similarities to estimates from other data sources are instructive.

In contrast to Akan et al. (2024), I find that the rise in distance among new hires has not been monotonic: while post-2020 hires initially exhibited longer commutes, this trend moderates over time, suggesting a rebalancing of work arrangements. Finally, the magnitudes of my estimates differ from theirs non-trivially. For instance, I find that mean distances increased only about 10%, from a baseline of about 34 miles, rather than the near-doubling they find from a baseline of about 15 miles.

In tension with this paper and existing literature, Census data—such as the American Community Survey U.S. Census Bureau 2024—find that average commute times have remained relatively stable or even decreased since the pandemic. One plausible explanation for this discrepancy is an increase in remote workers who live further from their workplaces.

Overall, this paper contributes to a growing understanding of how remote work has redefined labor market geography, not only through long-term migration patterns but also through changes in firm hiring strategies, job tenure, and occupational mobility.

3.3 Data

This paper primarily uses the Longitudinal Employer Household Dynamics (LEHD) dataset in my analysis, which is restricted-use administrative linked employer-employee data from the U.S. Census Bureau. The LEHD covers 95% of private-sector employment in the US (Graham et al. 2022) and combines information from state unemployment insurance (UI) records with establishment-level data from the Quarterly Census of Employment and Wages (QCEW), enabling the construction of detailed individual-level labor market histories.

I use LEHD quarterly earnings data from 30 states [1]. Collectively, these states accounted for roughly 51% of the U.S. population as of the 2010 Census and provide substantial geographic diversity, including representation from each major census region. The available quarterly data span varying start dates by state and extend through the first quarter of 2024. These records include job-level information such as the location of the employing establishment, quarterly earnings, and industry codes.

In addition to employment records, the LEHD provides annual residential address data for individual workers, covering all 50 states and available from 1999 to 2022. These data are

[1] AL, AZ, CO, CT, DC, DE, HI, ID, IL, KS, LA, ME, MD, NE, NV, NJ, NM, ND, OH, OK, OR, PA, SC, TN, TX, UT, VA, WA, WI, WY. See Figure 1.1 for a map.

geocoded at the Census block level, enabling precise measurement of home locations. For the purpose of my analyses, I extract the latitude and longitude coordinates of both workers' residential locations and their primary employer establishments.

The LEHD also includes rich demographic information, such as race, ethnicity, gender, age, and country of birth, along with firm characteristics including industry (at the six-digit NAICS level), firm size, and total payroll.

To measure geographic distance between workers and their employers, I calculate the haversine distance—a standard method for computing great-circle distances between two points on the Earth's surface—using the residential and employer coordinates. This approach yields a consistent, high-resolution measure of the spatial separation between workers and their jobs, which is aggregated from the individual level into various summary statistics for use as the primary outcome variables throughout the paper. In particular, I use the mean, median, and percent of employees living 50 or more miles from their workplaces as various measures of the main results. For the mean calculations, I winsorize the distances at 250 miles (a distance roughly equivalent to four hours of driving) in order to address potential extreme outliers.

My sample consists of 2,065,000,000 person-quarter observations, 104,700,000 persons, and 6,314,000 firms[2], between 2016 Q1 and 2022 Q4.

3.4 Results

I present my empirical results in Figures 3.1 through 3.6. Together, these results reinforce and extend the emerging consensus that the pandemic catalyzed a durable, albeit evolving, shift in the geography of work.

A noteworthy feature of the data is the temporal mismatch in the granularity of residential and workplace information: while residential locations are reported annually, workplace locations are reported quarterly. This discrepancy introduces systematic seasonal variation into the measured distances, with increasing distances within each year as workers change workplace locations each quarter but appear to remain in the same residential location for the year. To address this, I collapse the data to the annual level by averaging quarterly distances within each year. Other potential smoothing options include computing moving averages over rolling quarters, or focusing on a specific quarter (such as Q2) for each year to maintain temporal consistency. Figures C.1 through C.7 in Appendix C show the raw, unsmoothed quarterly data.

[2]These figures are rounded to comply with Census disclosure requirements.

3.4.1 Increasing post-pandemic home-work distances

Figures 3.1 through 3.3 depict the evolution of work-home distance over time, using three distinct metrics: mean distance, median distance, and the share of workers residing 50 or more miles from their job site. Across all measures, there is a clear upward shift in work-home distance beginning in 2020Q2, coinciding with the onset of the COVID-19 pandemic. This pattern aligns with prior research, including Akan et al. (2024), and reinforces the view that remote and hybrid work arrangements enabled greater spatial flexibility in labor markets.

In comparison to Akan et al. (2024)'s work, however, I estimate non-trivially different magnitudes. My baseline distance estimates are higher, with pre-pandemic work-home mean distances around 34 miles rather than 15, medians of around 10.5 miles rather than 6.5, and the percentage of employees living 50 or more miles away from work around 16% rather than 4%.

Furthermore, perhaps due to higher baseline estimates, I find smaller percentage post-pandemic increases in work-home distances as well, finding a roughly 10% increase in mean distance and median distance, respectively, as opposed to near doublings.

There are several possible explanations for this. First, my sample is far more comprehensive in terms of the number of individuals and firms, with over 100 million and 6 million respectively. This arguably represents a more nationally representative view of commuting patterns across a broader swath of the U.S. workforce. Adjusting for demographic representation alone may not fully capture underlying relevant factors, such as housing differences or regional infrastructure, contributing to work-home distance.

Second, however, my sample is limited in its geographic range, excluding large states such as California and New York. Given the concentration of major metropolitan areas in these states, their exclusion may lead to upward-biased estimates of work-home distances. Regardless of the underlying reasons for differing magnitudes across the literature, these estimates represent a useful empirical benchmark for comparison.

Third, although my sample offers a longer pre-pandemic time series, it is less current than some other sources, with residential data available only through 2022. A more recent time series could either reinforce or challenge the trends observed here.

Finally, a possible explanation lies the way the LEHD treats multi-unit firms: for workers who work in firms that have more than one establishment within a state, the LEHD imputes the particular establishment in which the employee works, based on the residential location of the employee [3]. However, this imputation plausibly biases my estimates downward rather

[3]In undisclosed results, I use a sample of firms that have only a single establishment within a state to check robustness to this issue.

than upward, as workers are more likely to be imputed to establishments closer to their homes than further away.

3.4.2 Hiring cohort

Figures 3.4 and 3.5 explore how these trends vary by hiring cohort. Specifically, I distinguish between workers hired before and after 2020Q2, which I define as the start of the pandemic period. Workers hired during or after the pandemic exhibit significantly longer work-home distances, with averages of roughly 50 miles in comparison to pre-pandemic averages of roughly 35 miles (Figure 3.4), suggesting that employers expanded their geographic recruitment ranges in the wake of remote work adoption.

Notably, however, the average distance among post-pandemic hires declines gradually over subsequent quarters, from about 51 miles in 2020 to 46 miles in 2022. This attenuation is suggestive of a partial re-convergence in workplace-residence proximity, possibly due to evolving employer expectations, hybrid work arrangements, or formal return-to-office policies. This trajectory—an initial expansion followed by gradual contraction—has not been well-documented in existing literature and provides a novel contribution.

3.4.3 Industry

Figure 3.5 presents trends disaggregated by industry, based on two-digit NAICS codes. I group industries into three categories based on their expected remote work feasibility and observed distance shifts: high-mobility (Information), medium-mobility (Professional, Finance, and Administrative Services), and low-mobility (all remaining industries). Unsurprisingly, the Information sector exhibits the most pronounced increase in work-home distance, increasing from a 17% pre-pandemic share of employees living 50 or more miles from work to about 22% in 2022, consistent with the sector's high suitability for remote work. Medium-mobility industries show moderate increases, while low-mobility sectors exhibit relatively stable patterns, reinforcing the view that occupational and industry characteristics significantly shape the spatial response to remote work capability.

3.4.4 Gender

Finally, Figure 3.6 examines heterogeneity by gender. The results show that men have consistently higher average commuting distances than women both before and after the pandemic. Nevertheless, both genders experienced a clear upward shift in distance in the post-pandemic period, with the share of employees living 50 or more miles from work in-

creasing from about 17.5% to 19% for men and 14.5% to 17% for women. This finding is consistent with the broader literature documenting gender differences in commuting patterns, but also highlights that both men and women participated in the post-pandemic geographic decoupling of home and workplace.

3.5 Conclusion

The COVID-19 pandemic initiated a structural shift in where and how Americans work, loosening the traditional geographic link between home and workplace. This paper provides new evidence on how the spatial distribution of U.S. workers and jobs has evolved in response, using detailed administrative data from the LEHD to track work-home distances before and after the onset of the pandemic.

I find that average commuting distances have increased since early 2020, driven largely by individuals hired after the pandemic began. This suggests that the hiring practices of firms have become more geographically expansive, likely reflecting greater flexibility in remote and hybrid work arrangements. However, this trend has not been uniformly sustained: among new hires, commuting distances began to decline after their initial post-pandemic peak, indicating that some of the early geographic dispersion may have moderated in the face of shifting organizational norms and return-to-office efforts.

Importantly, the growth in work-home distances is not evenly distributed across sectors or demographics. The most substantial changes are concentrated in industries with high remote work potential, such as Information and Professional Services, and overall distance increases are present across both male and female workers. These patterns point to persistent, but heterogeneous, shifts in the spatial configuration of employment relationships.

Taken together, these findings highlight a labor market that is still adapting to the expanded possibilities of remote work. While the initial disruption expanded the spatial boundaries of employment, ongoing adjustments suggest a dynamic process of equilibrium-seeking among workers and firms. As employers continue to refine their remote work policies and workers respond to new opportunities and constraints, the geography of work in the United States will likely remain in flux for years to come.

3.6 Figures

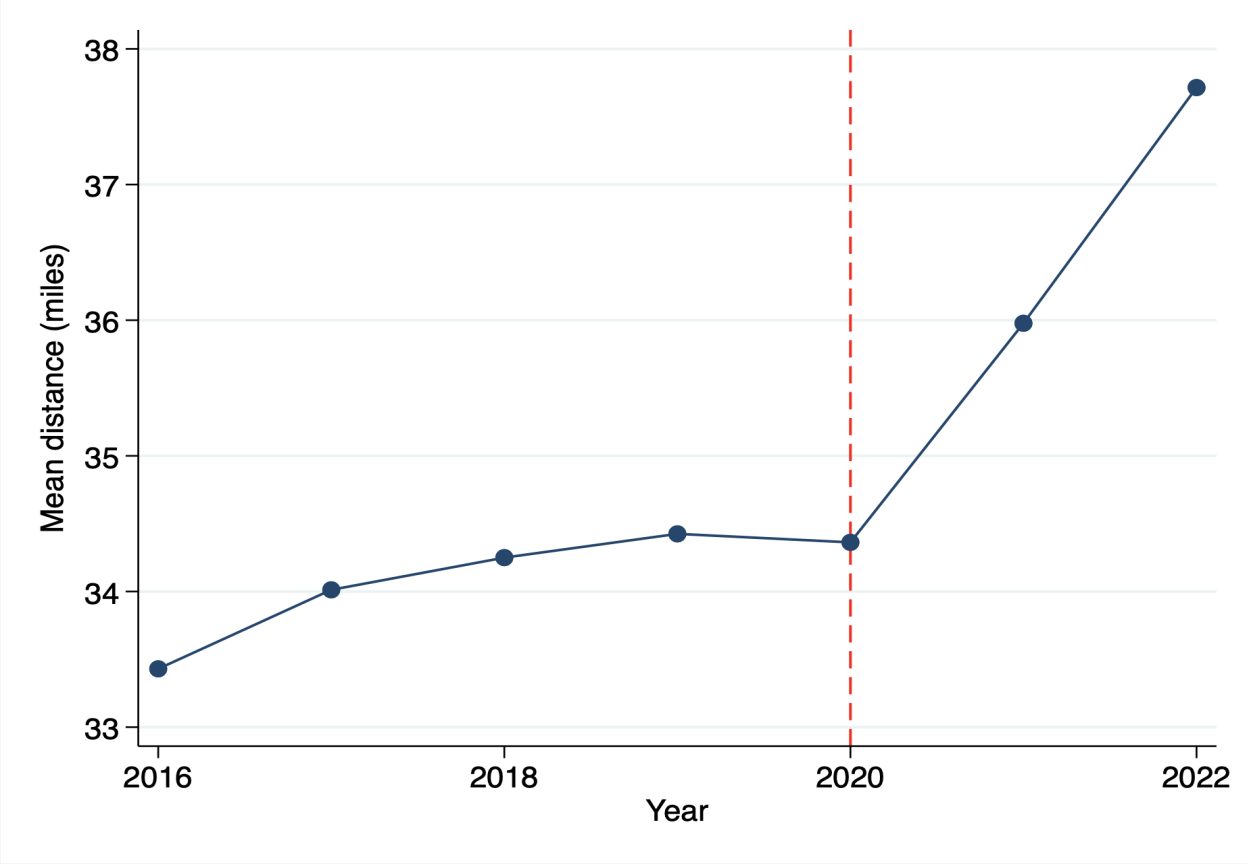


Figure 3.1: Mean Distances

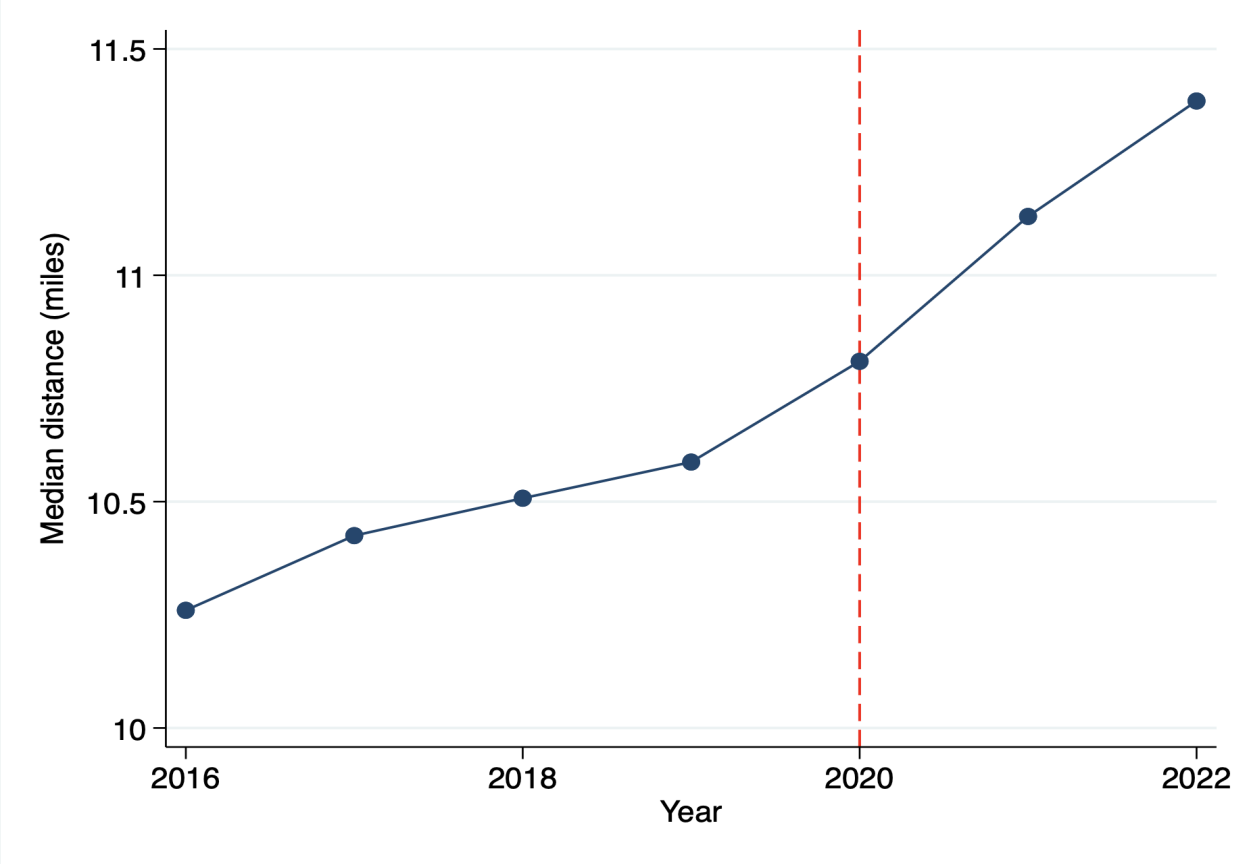


Figure 3.2: Median Distances

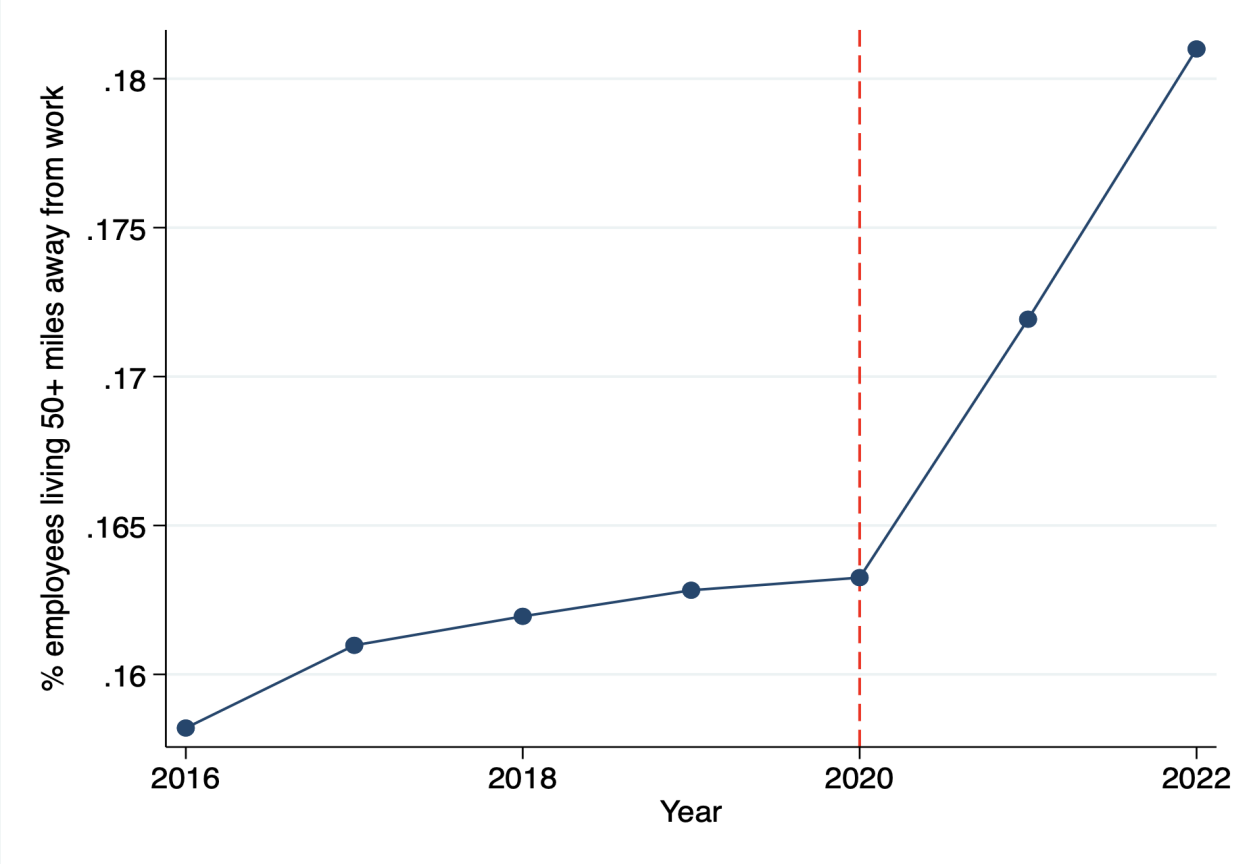


Figure 3.3: Percentage of Employees Living 50+ miles from Work

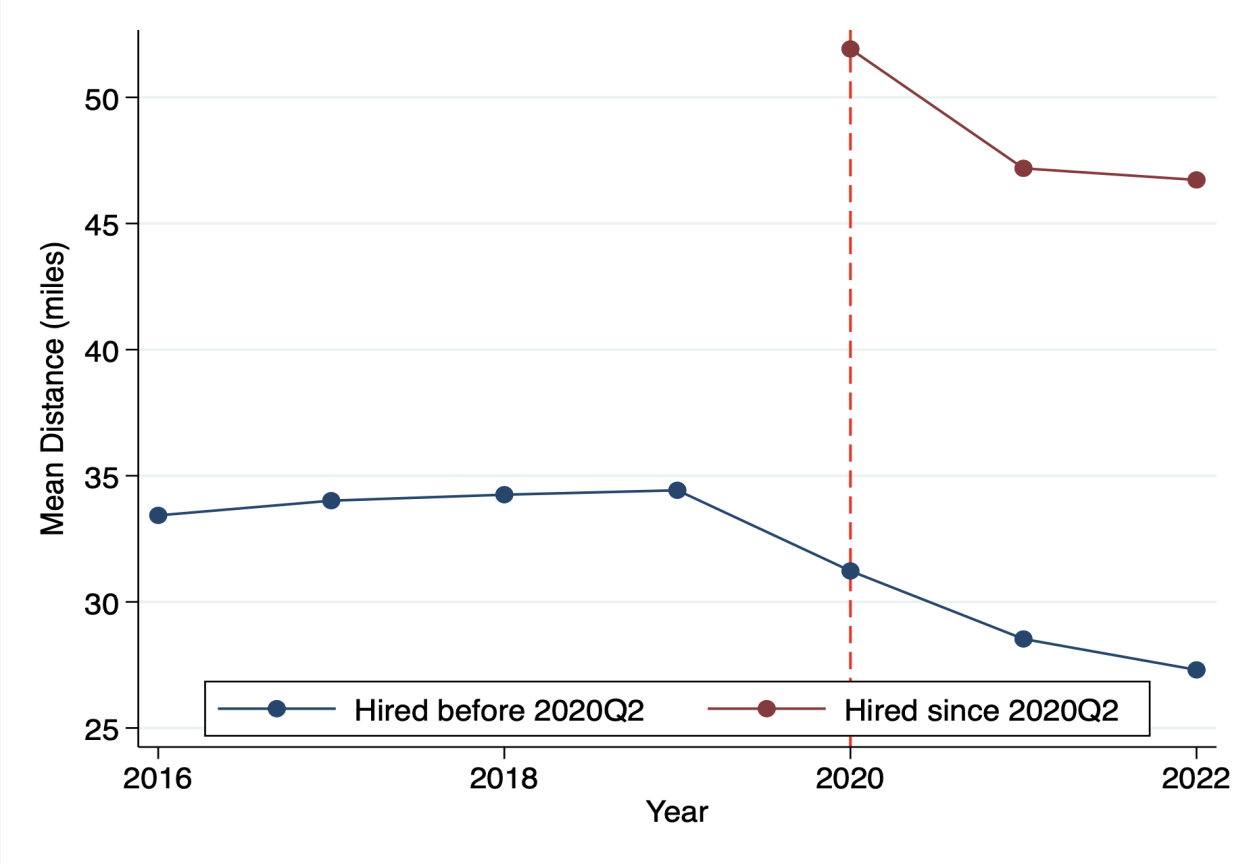


Figure 3.4: Average Distance, by hire cohort

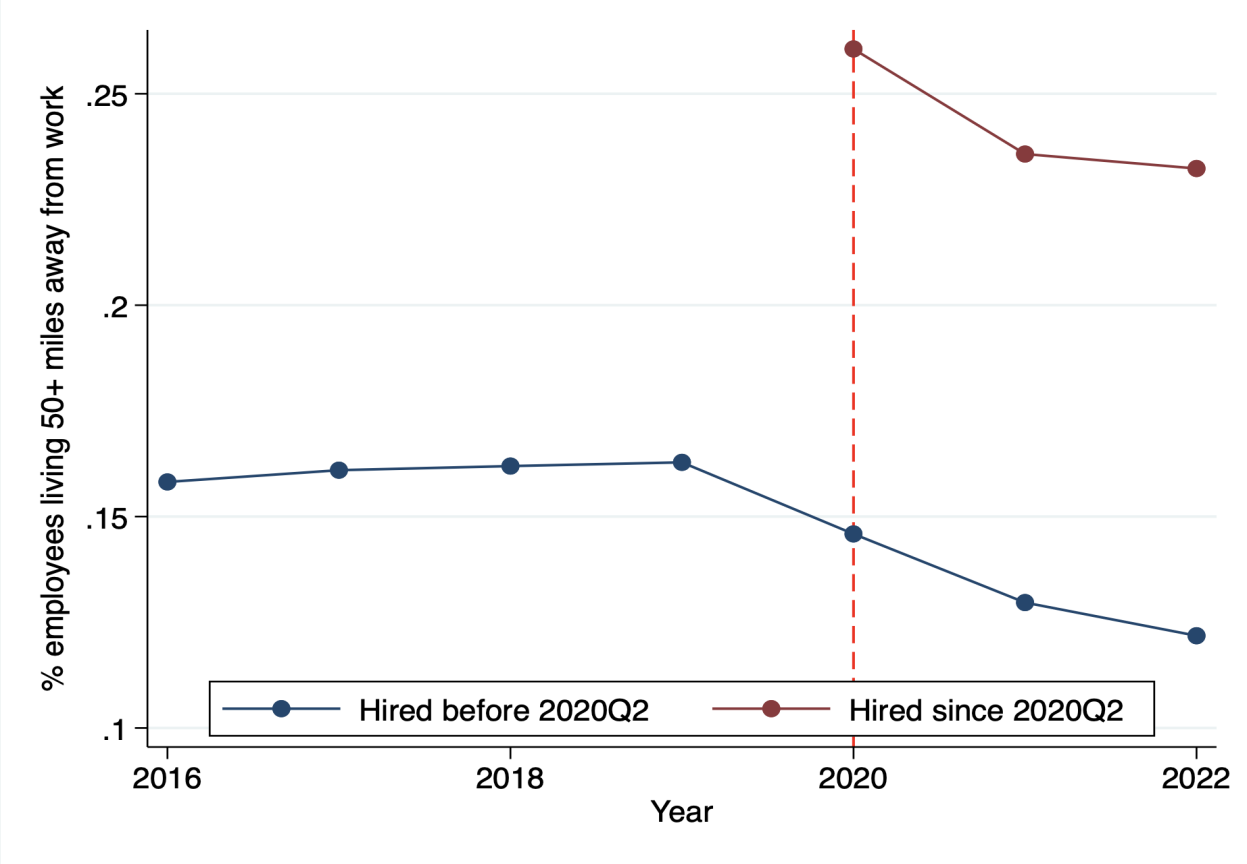


Figure 3.5: Percentage of Employees Living 50+ miles from Work, by hire cohort

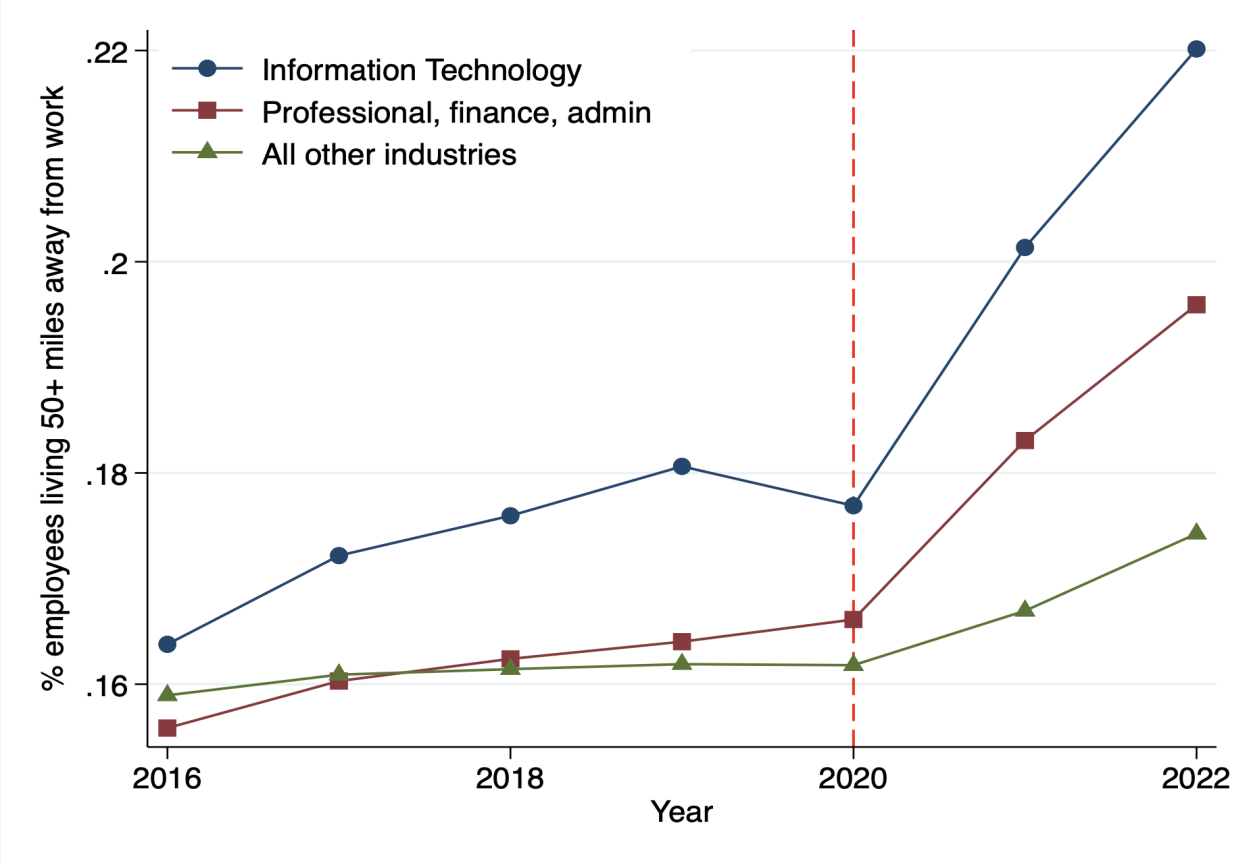


Figure 3.6: Percentage of Employees Living 50+ miles from Work, by Industry

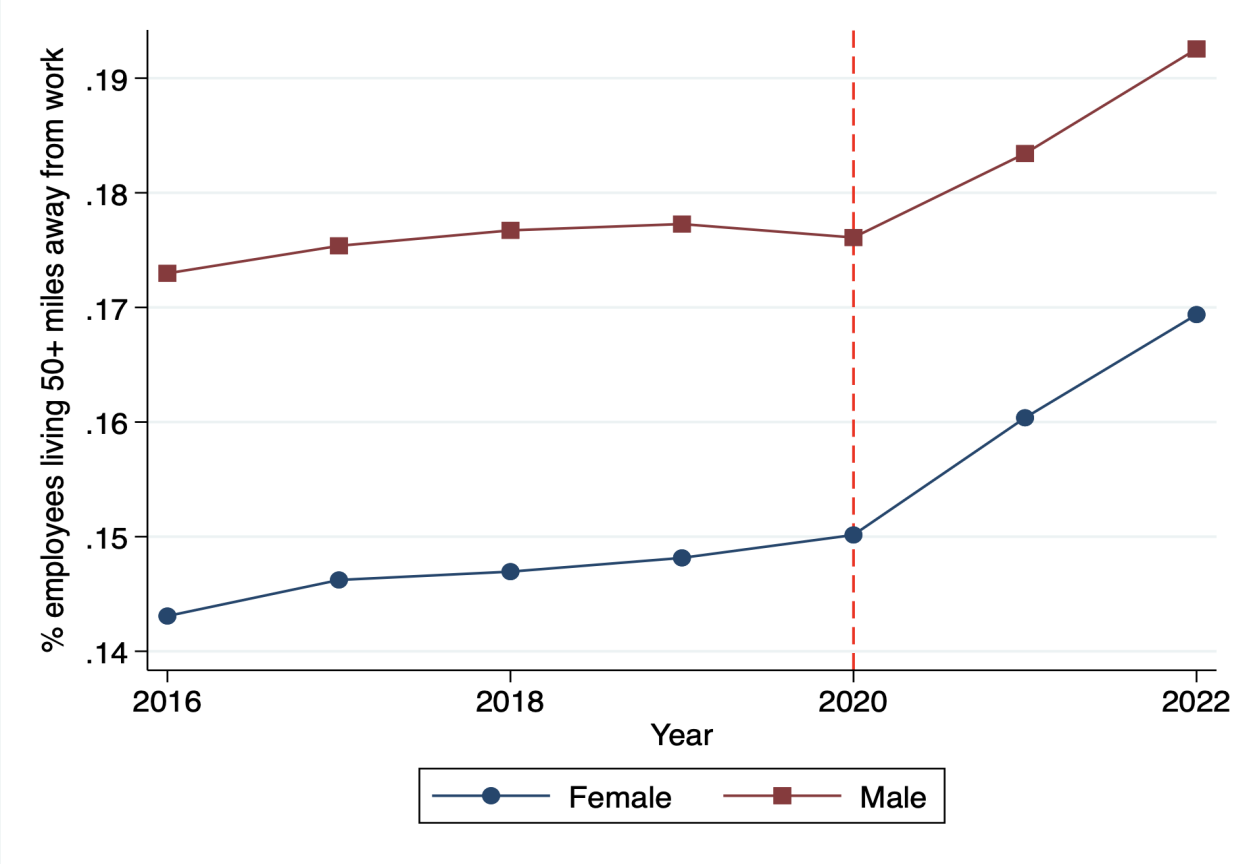


Figure 3.7: Percentage of Employees Living 50+ miles from Work, by Gender

APPENDIX A

Appendix to Chapter 1

A.1 Data Construction

I use the LEHD 2021 snapshot throughout. The main sample is constructed as follows:

Starting with the Residence Candidates File (RCF) from the Individual Characteristics File (ICF) residential geography files, I keep observations where workers live in one of the 50 states or the District of Columbia. I map counties to commuting zones following Autor and Dorn (2013). I then keep workers who changed their commuting zone (CZ) of residence between the two years of 2014 and 2019, where residential information in both years is non-missing. I further restrict the sample to prime-age adult workers workers aged (25-54) in 2014, where age is defined as of December 31 of that year. The lower bound is to address concerns about workers migrating for education-related reasons. The upper age bound is to address concerns about migration corresponding to retirement-related decisions.

Using the ECF SEIN for all 30 states available in this project, in my main sample, I keep firms that satisfy the following restrictions in 2014Q1: have between 2 and 1000 employees using the SEINSIZE_M variable, and have a single establishment within a state (i.e. have no more than one SEINUNIT per SEIN). I keep all firms in 2019 Q1, so that destination firms are unrestricted. In later samples I include firms regardless of number of establishments, or restrict to the manufacturing sector (NAICS 31-33).

Using the EHF for the 30 states available in this project, I merge with the workers and firms that satisfy my sample restrictions above, which results in a linked employer-employee sample that satisfy the above restrictions. I further restrict this sample to include only job observations in which a worker makes at least $ 15,000 a year, which is roughly full-time annual earnings at the federal minimum wage, and worker's highest paid job across all jobs they may hold in a particular year.

I then draw a 90% random sample of individuals which satisfy the sample restriction criteria for Census disclosure review purposes.

I then construct reference groups characterized jointly by 6-digit NAICS industry and commuting zone. I drop groups where only one firm comprises the group (i.e. there is only one firm in a particular industry and CZ). Workers in such groups will not have a non-coworker comparison group and thus will not contribute variation to the main regression. I also drop groups in which the number of workers is only 1 or 2 for similar reasons.

Within every industry by commuting zone reference group, I form pairwise dyads between all members of the group. This is done in order to implement the dyadic regression design, where the unit of observation is a pair of workers. This implies each worker will appear in multiple observations as a member of a different pair. I further restrict this set of pairs to exclude pairs of workers who live in the same address (given by the RCF's MAFID variable) in 2014.

In mapping state-level firms to national firms, I use the ECF T26 data, which contain data on SEIN-EIN-FIRMID combinations at the quarterly level. I interpolate the national firm identifier FIRMID for quarters in which it is missing if the same SEIN-EIN combination maps to a unique FIRMID in other quarters. I use FIRMID where it is available, and use EIN for quarters in which FIRMID is missing.

A.2 Appendix Tables

Table A.1: Effect of *coworker* on *comove*: Heterogeneous specification

coworker	0.012***
	(0.002)
coworker * both white	0.0033***
	(0.001)
coworker * both black	0.0024
	(0.002)
coworker * both asian	-0.0082***
	(0.002)
coworker * both hispanic	0.0054***
	(0.002)
coworker * both male	-0.0017**
	(0.001)
coworker * both female	-0.00029
	(0.001)
coworker * both native	0.000041
	(0.001)
coworker * both immigrant	0.0050**
	(0.002)
coworker * both 25-34	-0.00078
	(0.001)
coworker * both 35-44	0.0015**
	(0.001)
coworker * both 45-54	0.0072**
	(0.003
coworker * both wage $\leq$ 50th pctile	-0.0013
	(0.001)
coworker * both wage 50-90th pctile	0.0015**
	(0.001)
coworker * both wage > 90th pctile	0.036**
	(0.01)
N (pairs)	349,600,000
N (individuals)	1,635,000
Industry-CZ FE	Yes

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, have changed residence across commuting zones between 2014 and 2019, and work at firms with one establishment in their state in 2014. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The variable *coworker* is an indicator equal to one if both workers in the pair work in the same firm in 2014. The variable "coworker * both [demographic characteristic]" is an indicator equal to one if both workers share that demographic characteristic, interacted with the *coworker* variable. The regression also includes, but the table does not report, non-interacted indicators for each demographic variable. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table A.2: Effect of *coworker* on *comove*: Single and multi-establishment firms

Sample	Single and multi-estab firms
	$comove_{ij}$
$coworker_{ij}$	0.0154^{***}
SE	(0.0009)
N (pairs)	595,300,000
N (individuals)	2,429,000
Industry-CZ FE	Yes
$E[comove \mid coworker = 0]$	0.044
Effect size	35.4%

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, and have change residence across commuting zones between 2014 and 2019. Firms with any number of establishments are included. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

Table A.3: Effect of *coworker* on *comove*: Movers and nonmovers, manufacturing

Sample	Movers and Non-movers
	$comove_{ij}$
$coworker_{ij}$	0.00038***
SE	(0.00004)
N (pairs)	877,500,000
N (individuals)	2,163,000
Industry-CZ FE	Yes
$E[comove \mid coworker = 0]$	0.00042
Effect size	90.3%

Standard errors clustered at 6-digit industry-CZ level.

* p<0.10, ** p<0.05, *** p<0.01

The table reports the results of a regression where each observation is a pair of workers, each of whom is age 25-54, earn at least $15,000 annually, and work at manufacturing firms (NAICS 31-33) with one establishment in their state in 2014. Workers are included regardless of migration status. Both members of the pair work in the same 6-digit NAICS industry and commuting zone in 2014, and pairs in which both members live at the same address are excluded. The dependent variable is an indicator equal to one if both workers in the pair move to the same commuting zone by 2019, and zero otherwise. The independent variable is an indicator equal to one if both workers in the pair work in the same firm in 2014. Industry by commuting zone fixed effects are included. Standard errors are clustered at the 6-digit industry-CZ level. Sample sizes are rounded according to Census Bureau disclosure rules.

APPENDIX B

Appendix to Chapter 2

B.1 Appendix Tables and Figures

Figure B.1: Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation

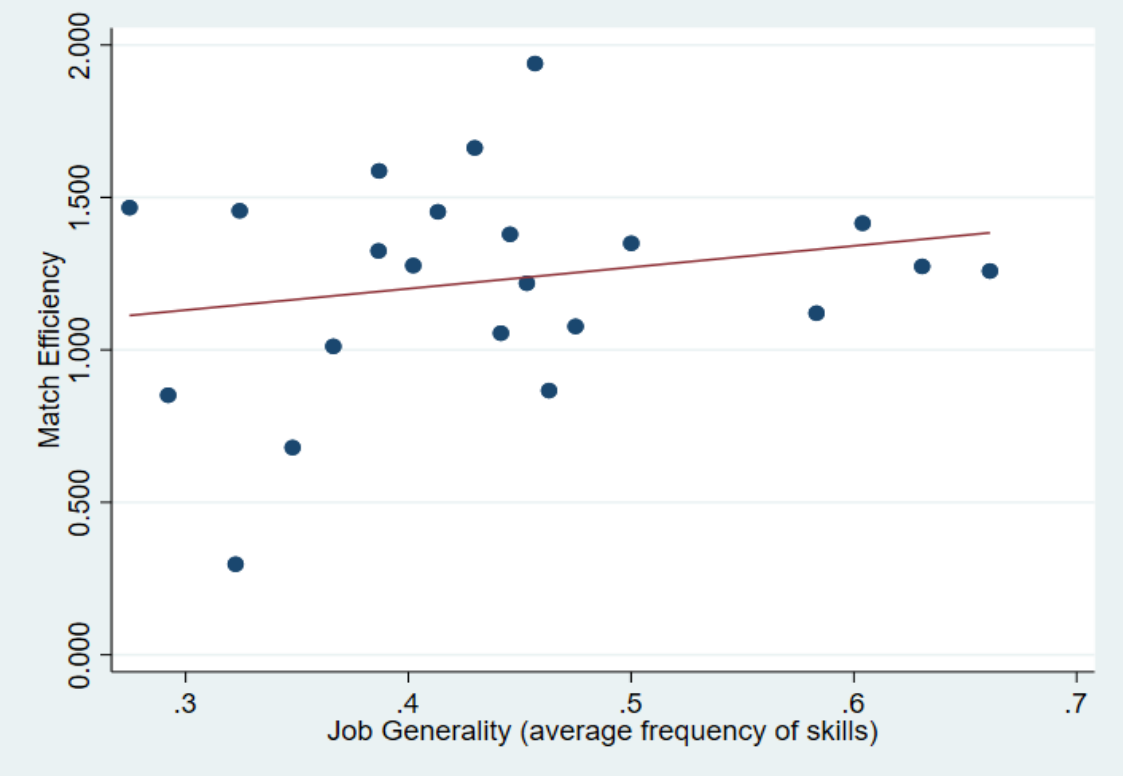


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-2 occupational level. Vacancies and unemployment are contemporaneous, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.2: Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation

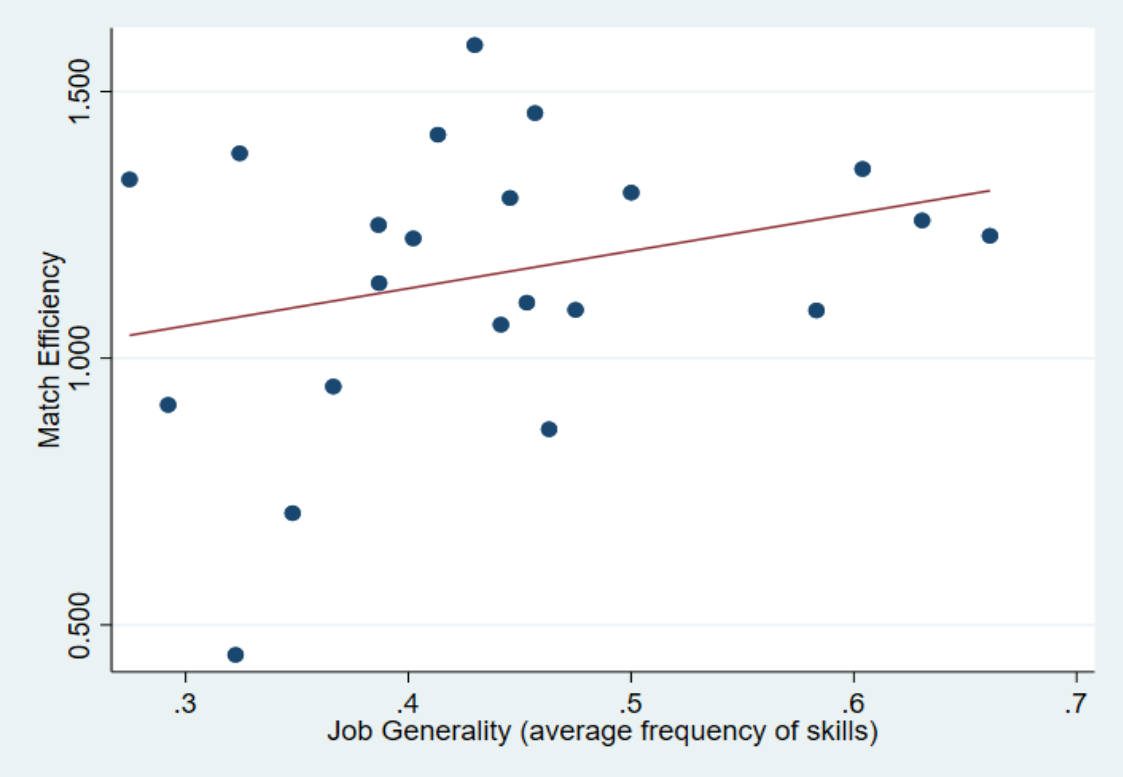


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-2 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Table B.1: Vacancy and Generality Variation Across States, 2010–2019

State	Total Vacancies	Avg Job Generality (Main)	Avg Job Generality (Alt.)	Avg # Skills
Alabama	26,135	0.00465	0.00482	7.80
Alaska	7,846	0.00423	0.00446	8.33
Arizona	53,782	0.00461	0.00488	9.42
Arkansas	14,626	0.00459	0.00476	7.78
California	274,779	0.00454	0.00487	10.18
Colorado	57,513	0.00468	0.00497	9.64
Connecticut	28,426	0.00452	0.00478	9.17
Delaware	7,905	0.00458	0.00487	9.47
District of Columbia	18,543	0.00433	0.00476	11.64
Florida	108,781	0.00477	0.00503	9.28
Georgia	69,749	0.00454	0.00483	9.59
Hawaii	8,658	0.00500	0.00522	8.66
Idaho	10,391	0.00466	0.00485	7.80
Illinois	90,409	0.00472	0.00502	9.65
Indiana	38,206	0.00477	0.00496	8.22
Iowa	27,320	0.00472	0.00489	7.85
Kansas	20,908	0.00467	0.00487	8.46
Kentucky	23,369	0.00476	0.00493	8.02
Louisiana	22,293	0.00482	0.00499	8.04
Maine	9,376	0.00471	0.00488	7.42
Maryland	46,032	0.00444	0.00475	10.12
Massachusetts	64,596	0.00461	0.00493	10.06
Michigan	70,364	0.00461	0.00484	8.75
Minnesota	51,894	0.00473	0.00498	9.02
Mississippi	12,168	0.00481	0.00497	7.65
Missouri	39,740	0.00468	0.00491	8.83
Montana	7,377	0.00456	0.00473	7.31
Nebraska	15,447	0.00462	0.00486	8.61
Nevada	20,929	0.00503	0.00526	8.72
New Hampshire	9,341	0.00453	0.00471	7.90
New Jersey	62,059	0.00463	0.00491	9.85
New Mexico	12,230	0.00449	0.00469	8.25
New York	116,212	0.00468	0.00498	10.07
North Carolina	61,410	0.00456	0.00484	9.63
North Dakota	6,445	0.00471	0.00487	7.21
Ohio	74,230	0.00465	0.00490	9.09
Oklahoma	20,678	0.00459	0.00479	8.24
Oregon	33,602	0.00458	0.00486	9.38
Pennsylvania	84,599	0.00467	0.00491	9.04
Rhode Island	7,980	0.00457	0.00477	8.14
South Carolina	24,829	0.00454	0.00473	8.28
South Dakota	7,512	0.00468	0.00484	7.28
Tennessee	37,967	0.00475	0.00498	8.81
Texas	169,122	0.00463	0.00491	9.62
Utah	22,457	0.00495	0.00520	9.13
Vermont	4,459	0.00444	0.00464	7.47
Virginia	69,292	0.00439	0.00469	10.45
Washington	57,538	0.00454	0.00485	9.98
West Virginia	8,190	0.00450	0.00465	7.39
Wisconsin	43,017	0.00465	0.00487	8.60
Wyoming	3,689	0.00446	0.00463	7.63

Figure B.3: Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation

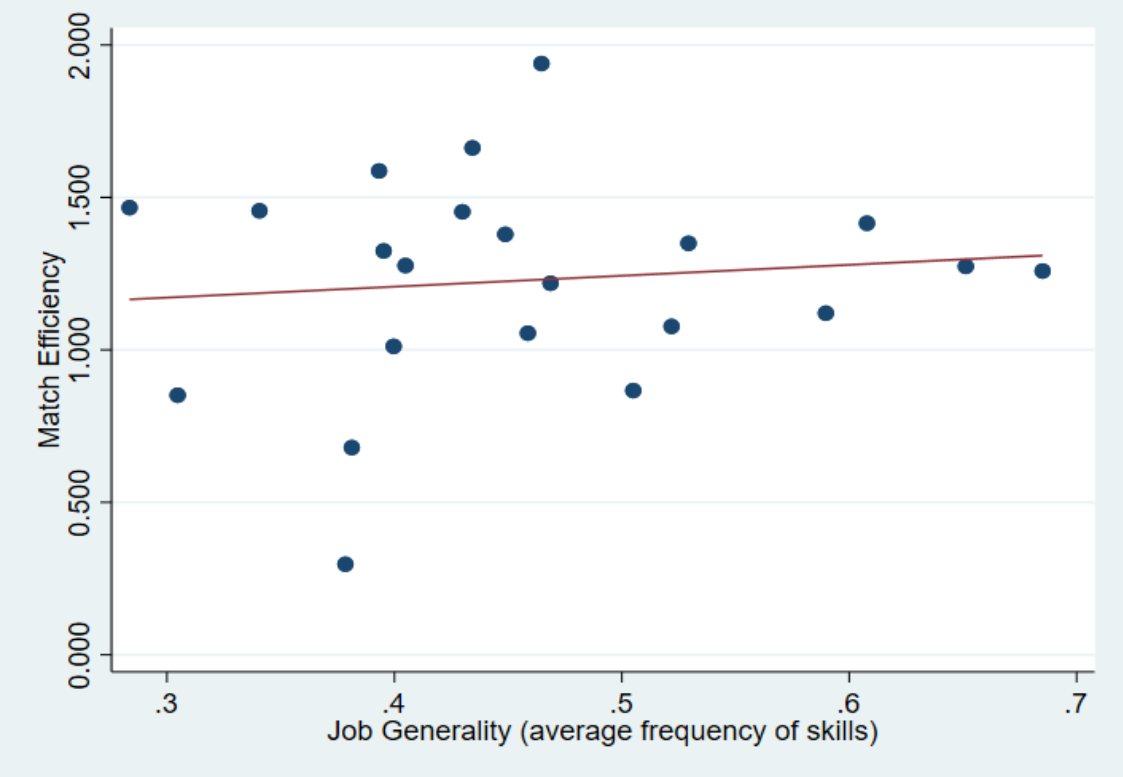


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-2 occupational level. Vacancies and unemployment are contemporaneous in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.4: Scatterplot of Matching Efficiency vs. Job Generality, SOC-2 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation

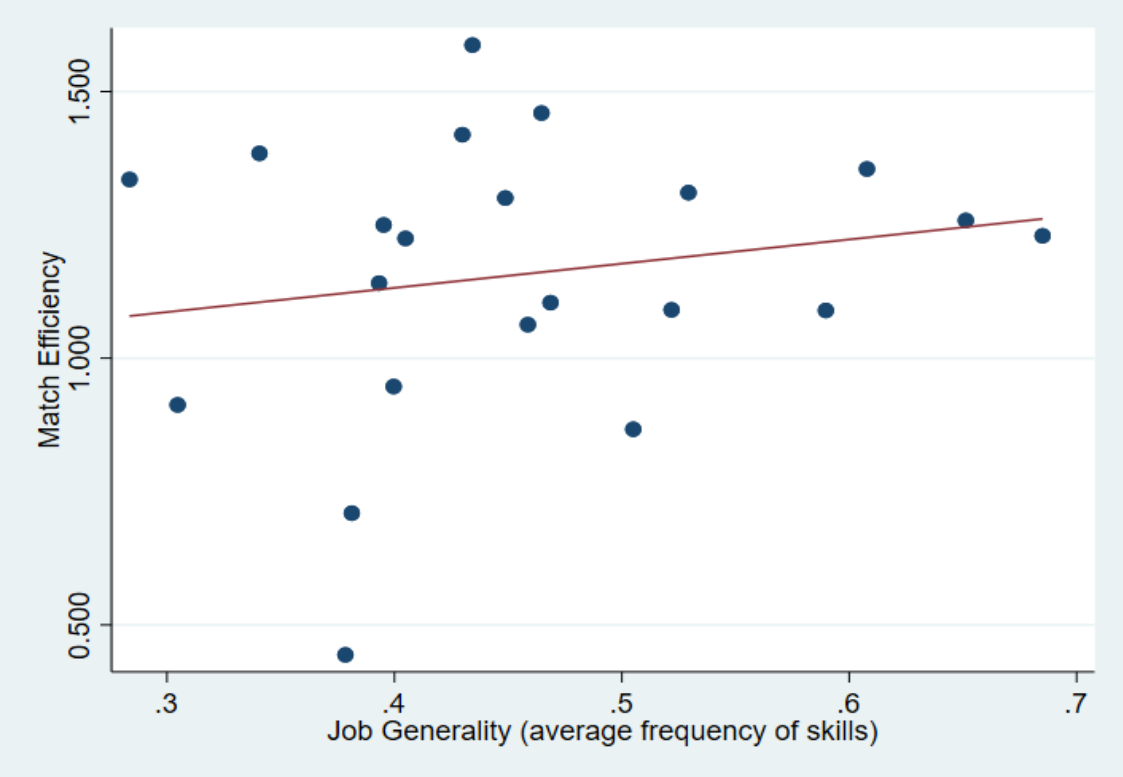


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-2 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.5: Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation

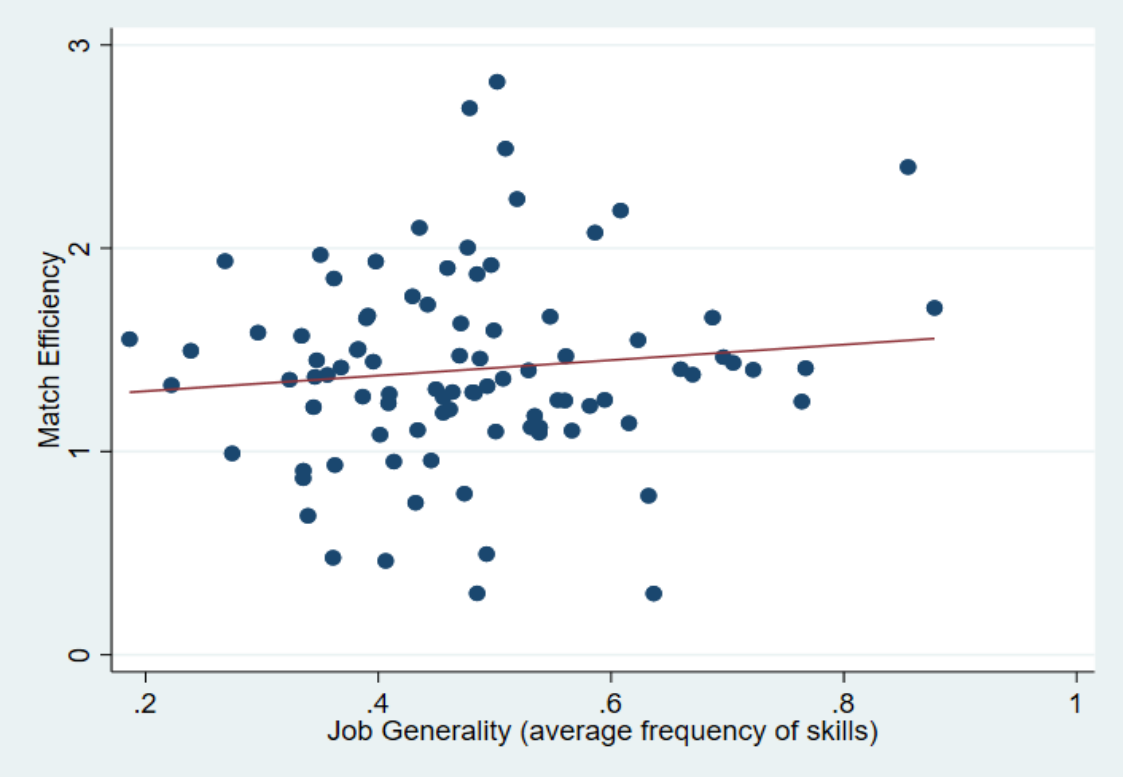


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-3 occupational level. Vacancies and unemployment are contemporaneous, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.6: Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation

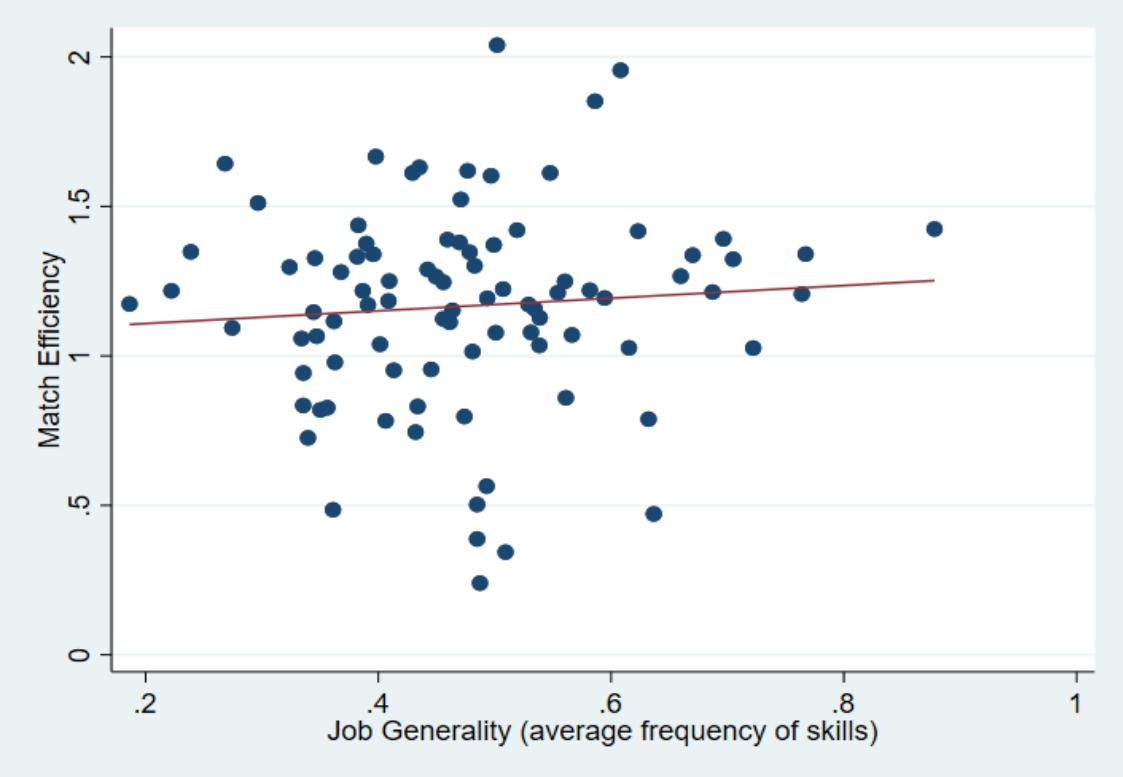


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-3 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.7: Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation

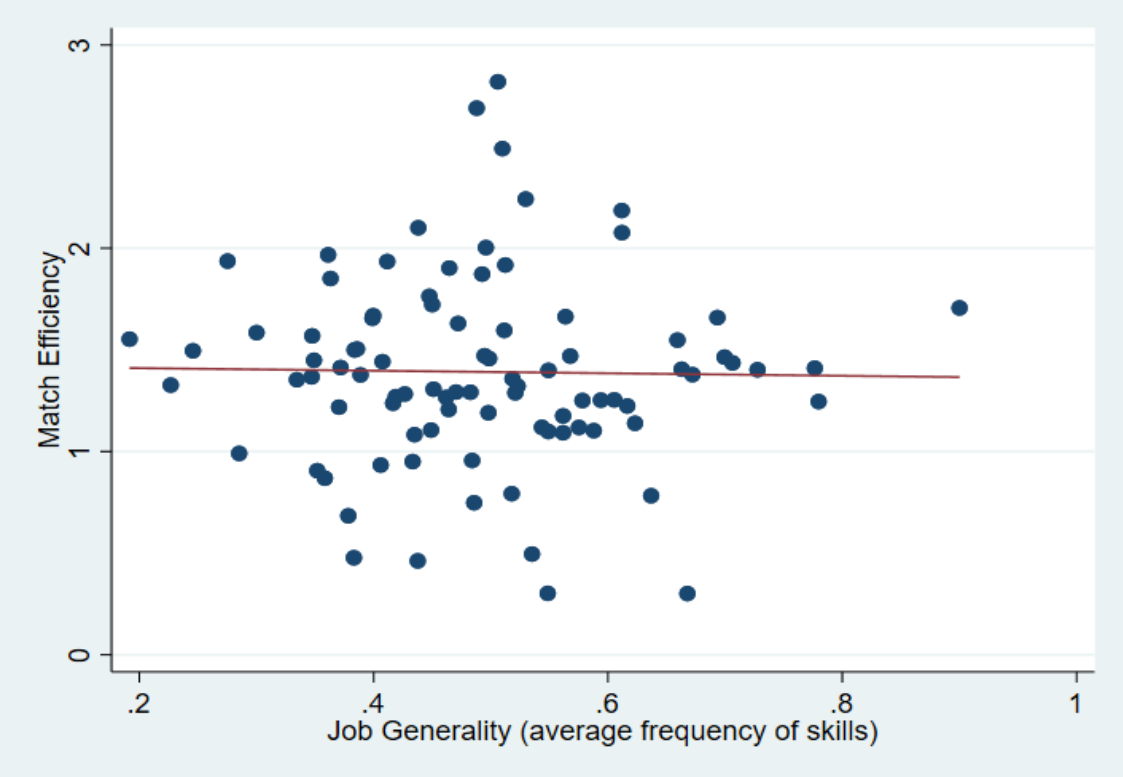


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-3 occupational level. Vacancies and unemployment are contemporaneous in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.8: Scatterplot of Matching Efficiency vs. Job Generality, SOC-3 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation

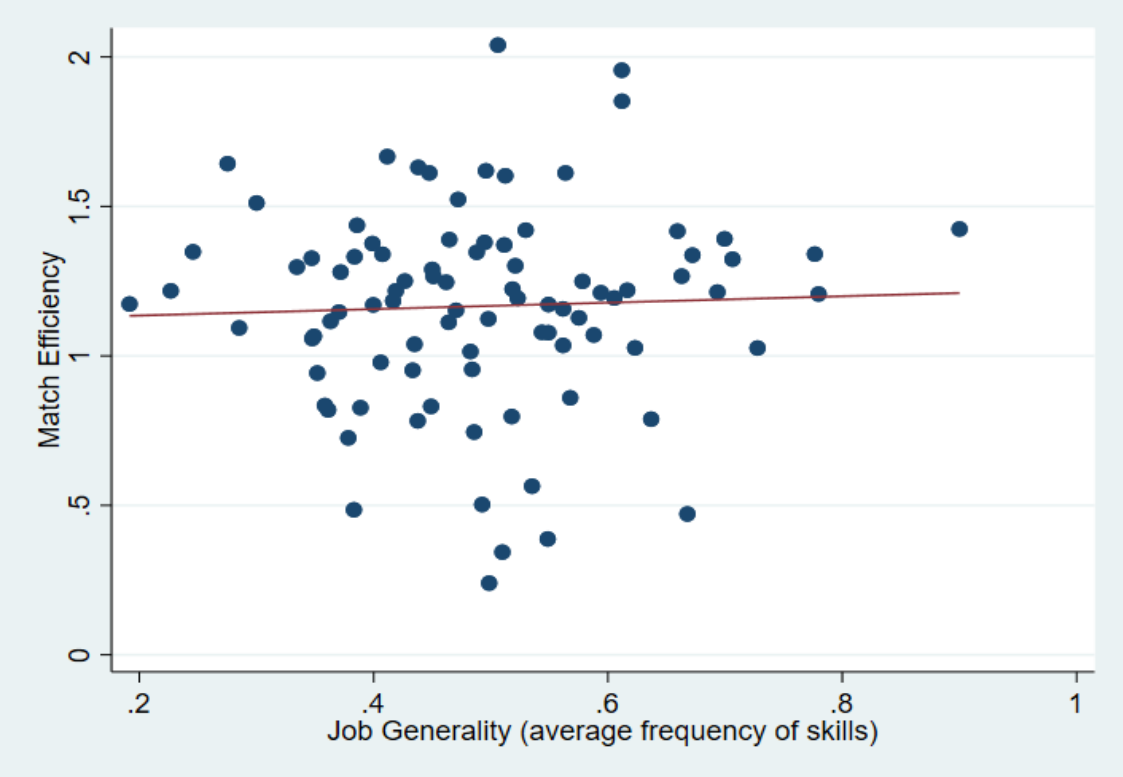


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-3 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.9: Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation

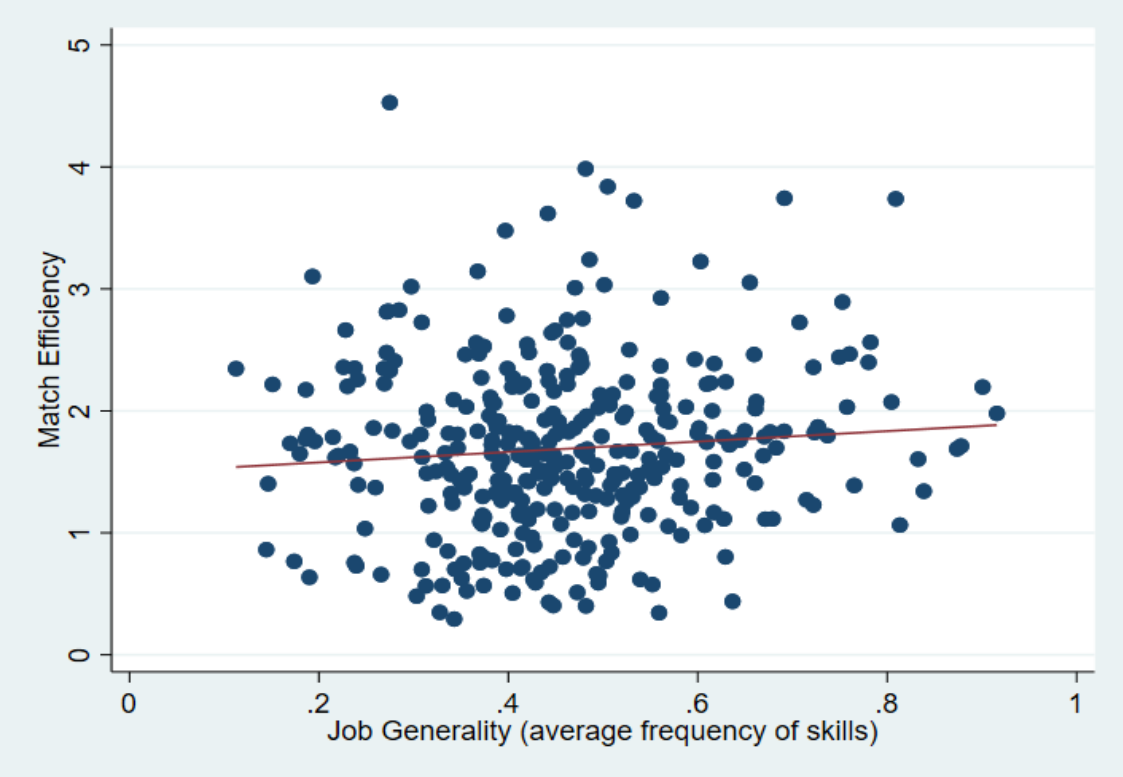


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-5 occupational level. Vacancies and unemployment are contemporaneous, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.10: Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation

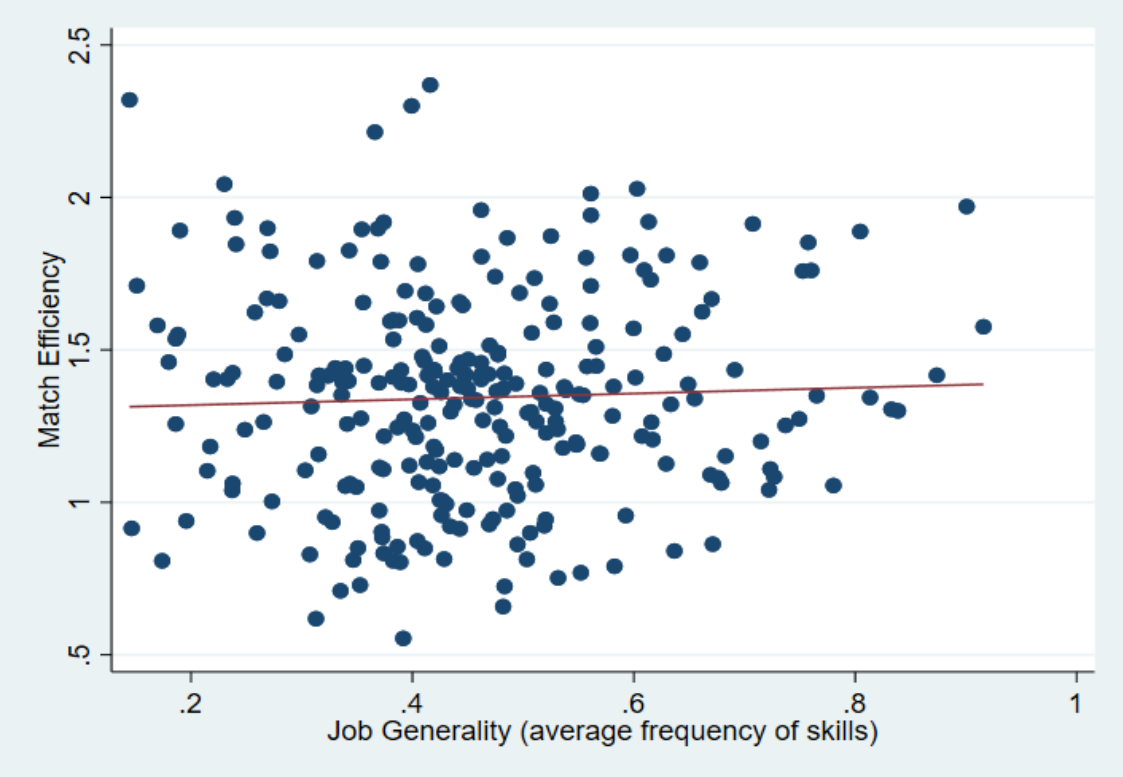


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-5 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.11: Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation

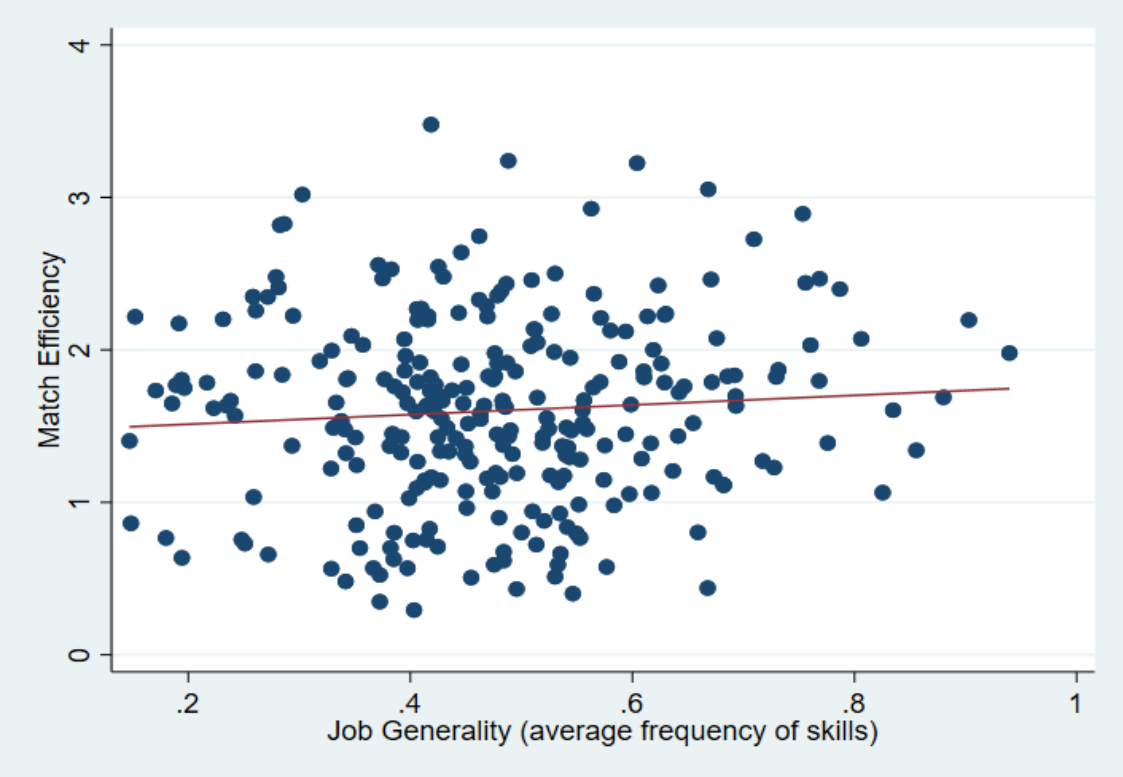


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-5 occupational level. Vacancies and unemployment are contemporaneous in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.12: Scatterplot of Matching Efficiency vs. Job Generality, SOC-5 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation

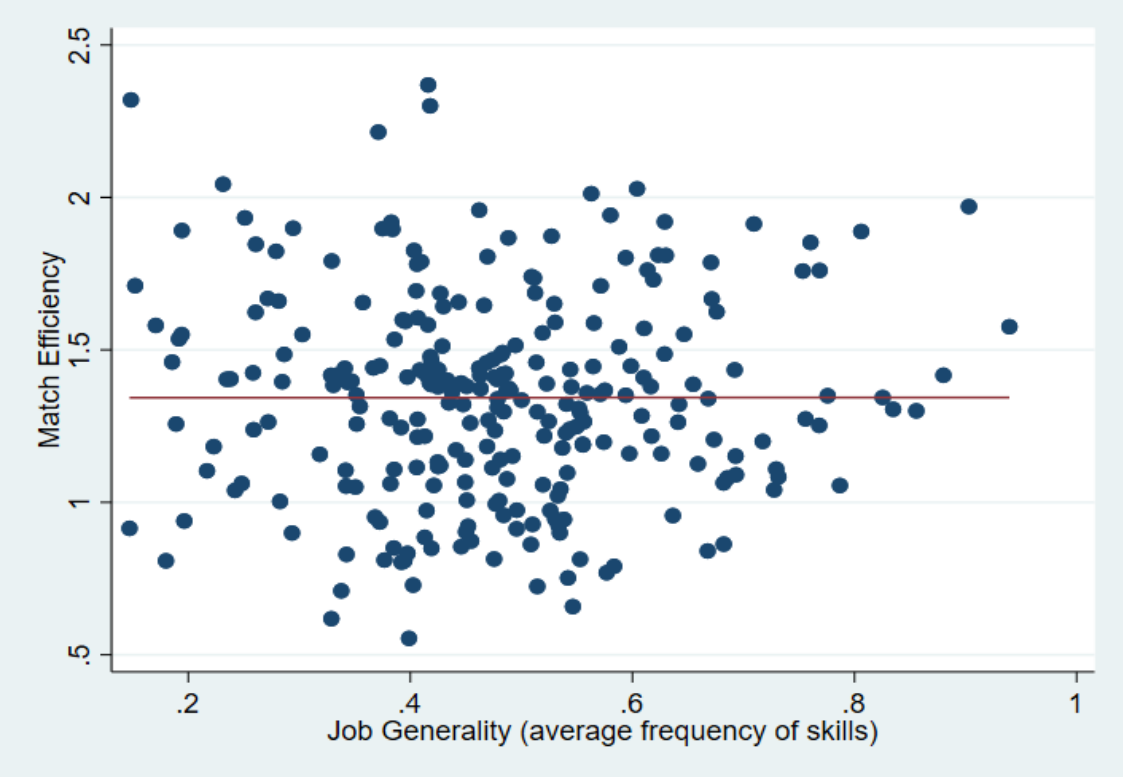


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-5 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.13: Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Main Generality Measure, Contemporaneous Matching Function Estimation

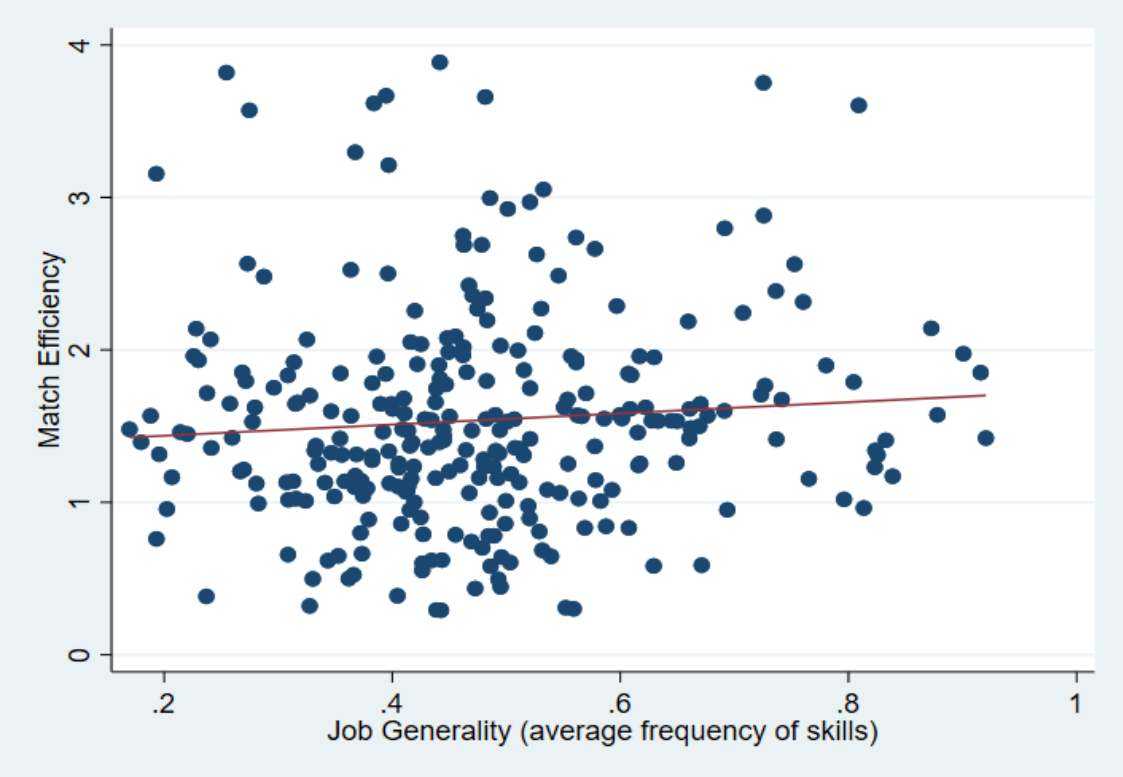


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-6 occupational level. Vacancies and unemployment are contemporaneous, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.14: Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Main Generality Measure, Lagged Matching Function Estimation

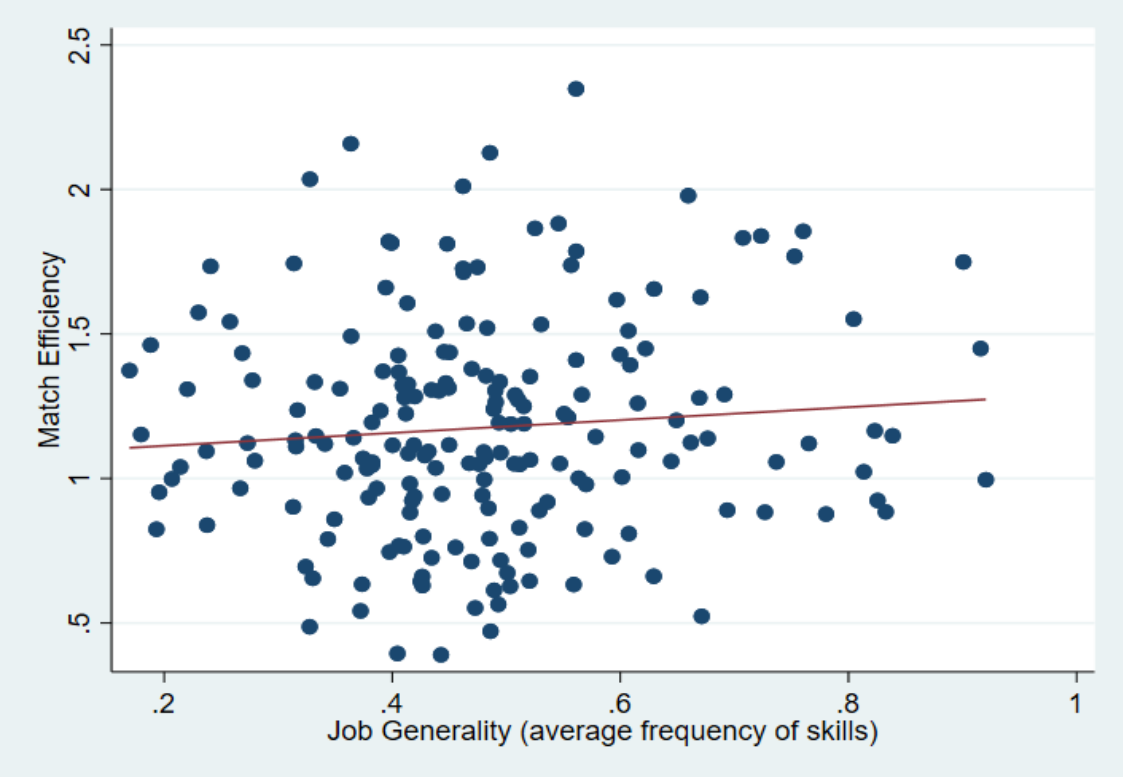


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-6 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the main measure using all skills. The red line shows a least-squares fit.

Figure B.15: Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Alternative Generality Measure, Contemporaneous Matching Function Estimation

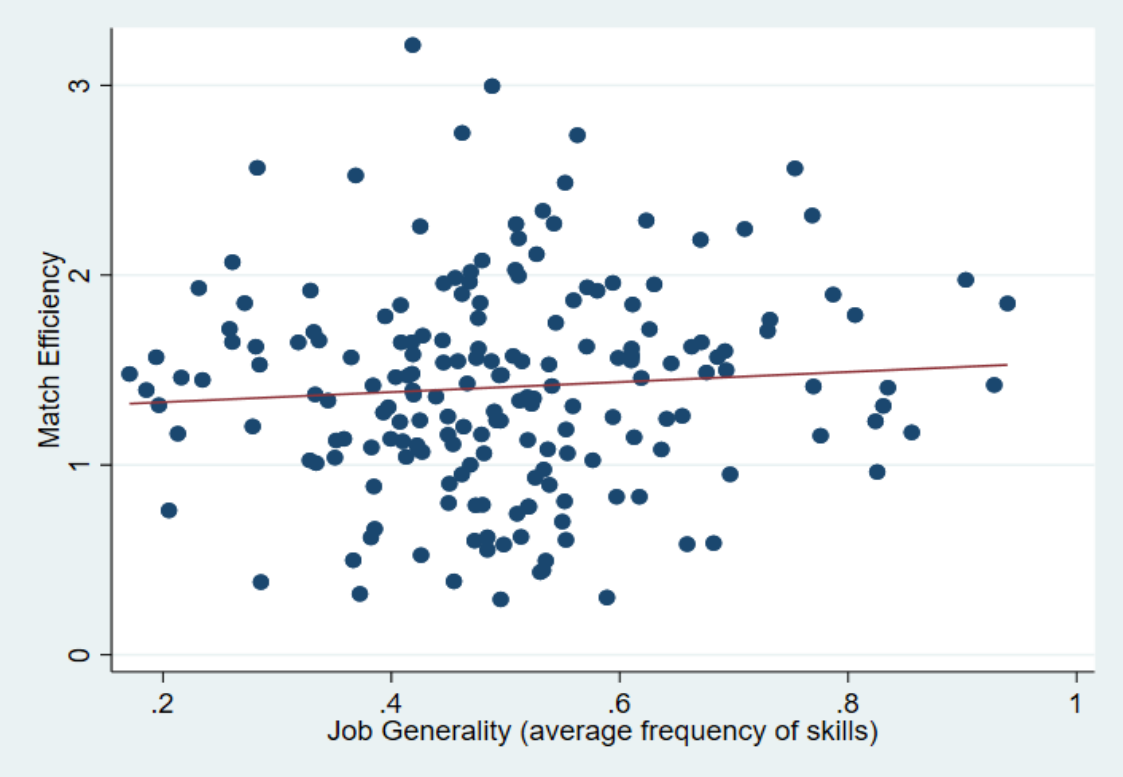


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-6 occupational level. Vacancies and unemployment are contemporaneous in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

Figure B.16: Scatterplot of Matching Efficiency vs. Job Generality, SOC-6 Occupational Level: Alternative Generality Measure, Lagged Matching Function Estimation

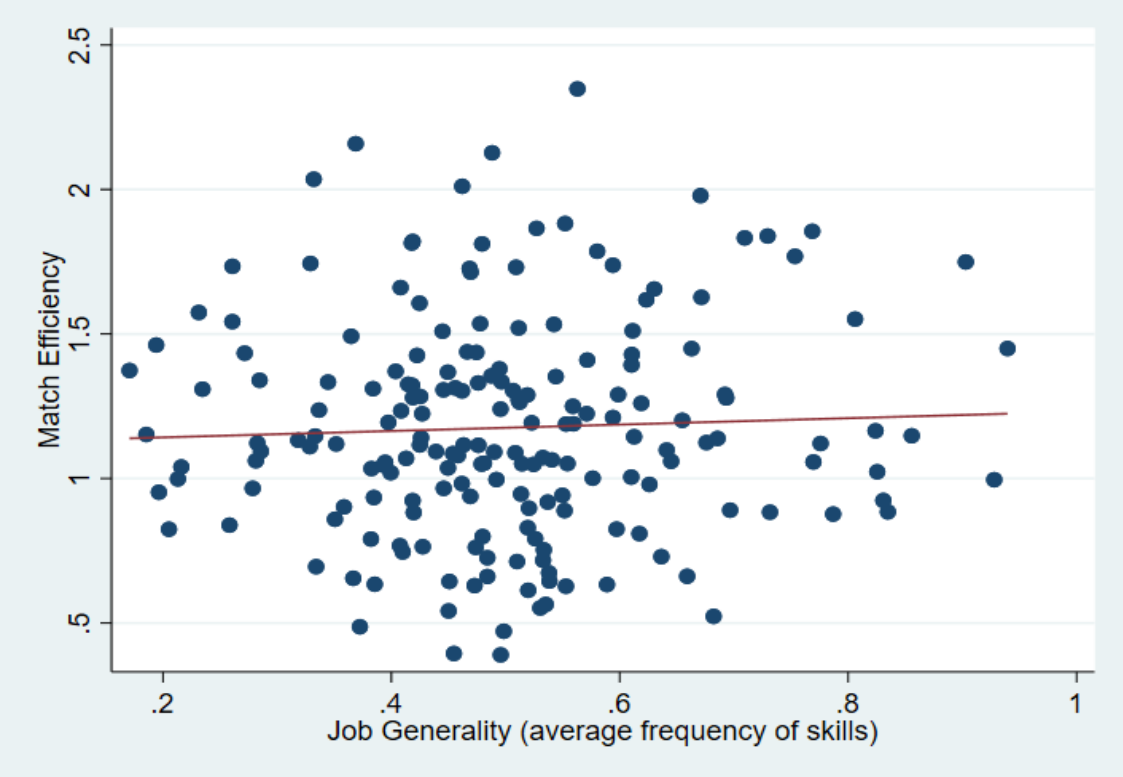


Figure plots matching efficiency, estimated from CPS hires and unemployment with Burning Glass vacancies, against job generality, at the SOC-6 occupational level. Vacancies and unemployment are lagged by one period in the matching function estimation, and the generality measure is the alternative measure using only the 15 most frequent skills per job. The red line shows a least-squares fit.

APPENDIX C

Appendix to Chapter 3

C.1 Appendix Figures - Quarterly

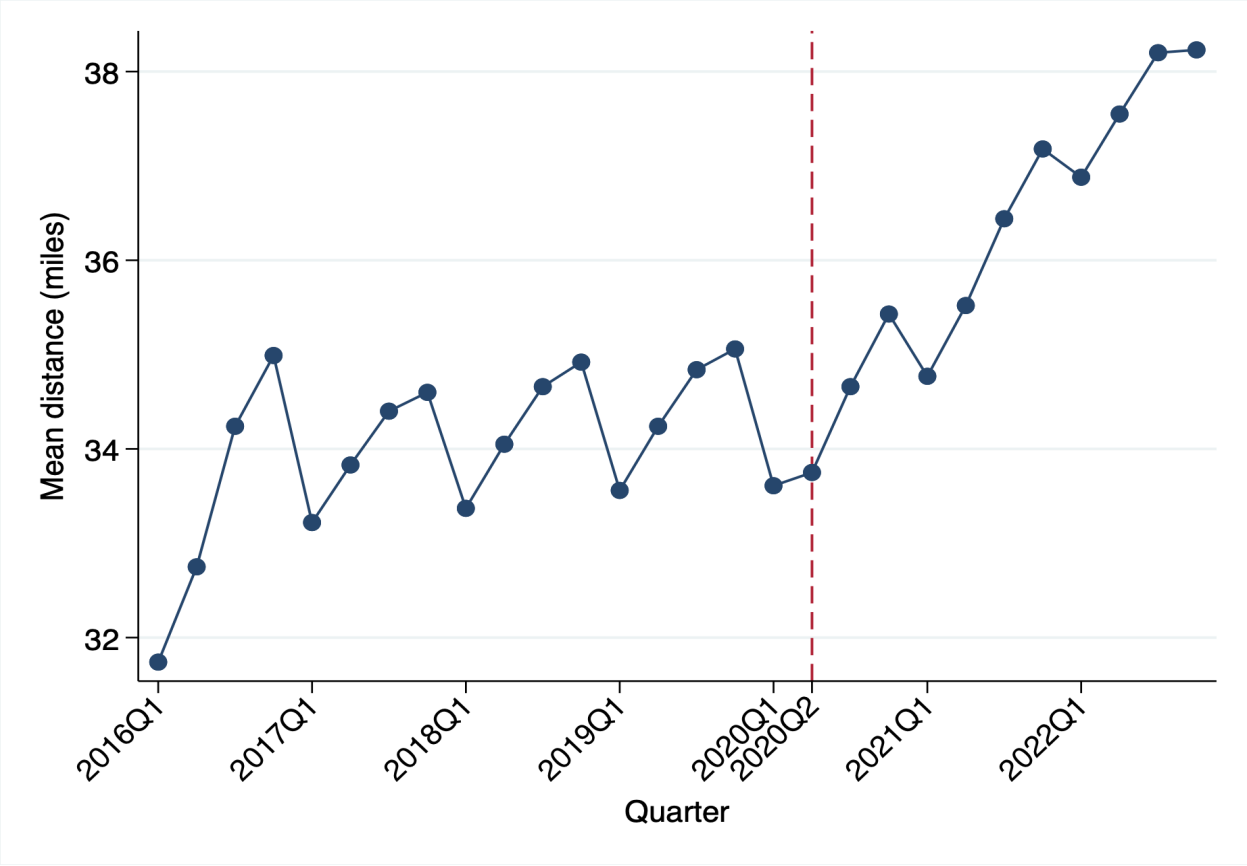


Figure C.1: Mean Distances, quarterly

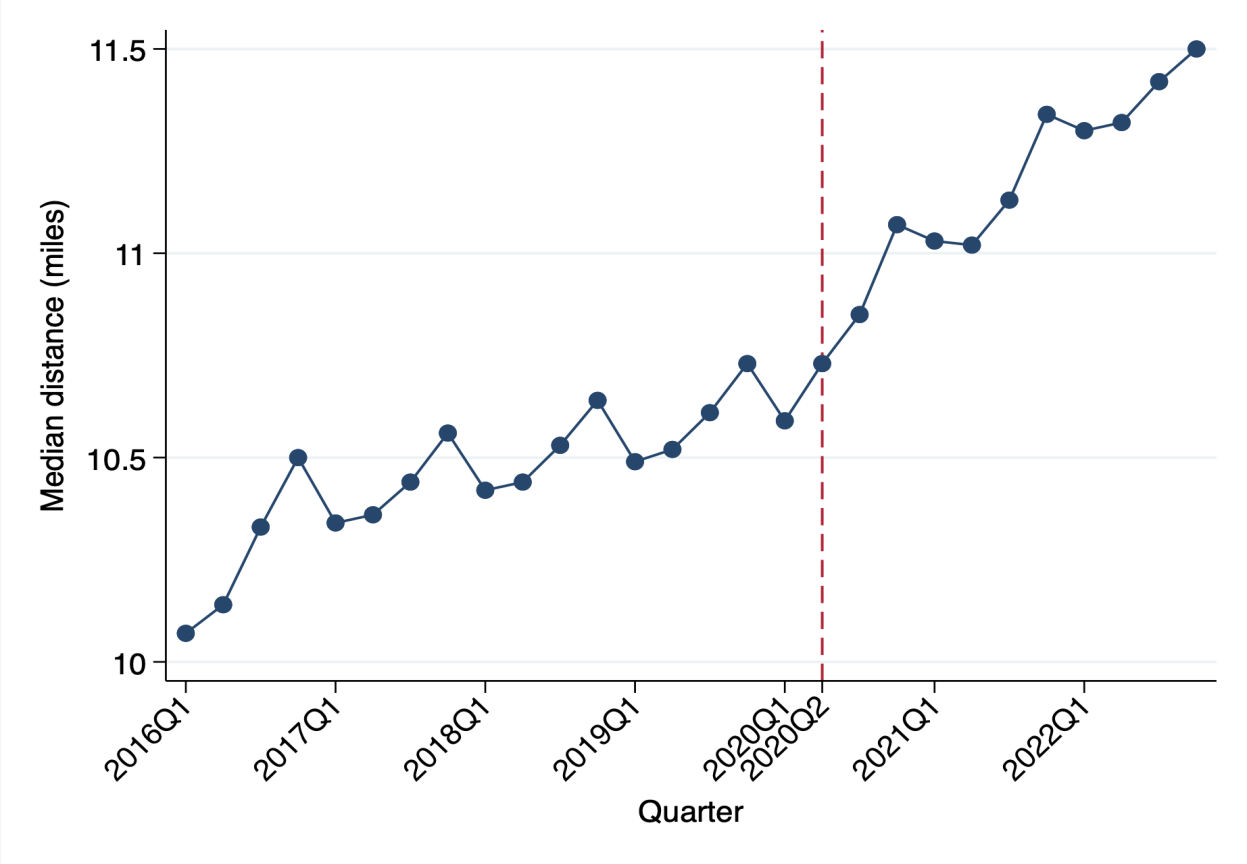


Figure C.2: Median Distances, quarterly

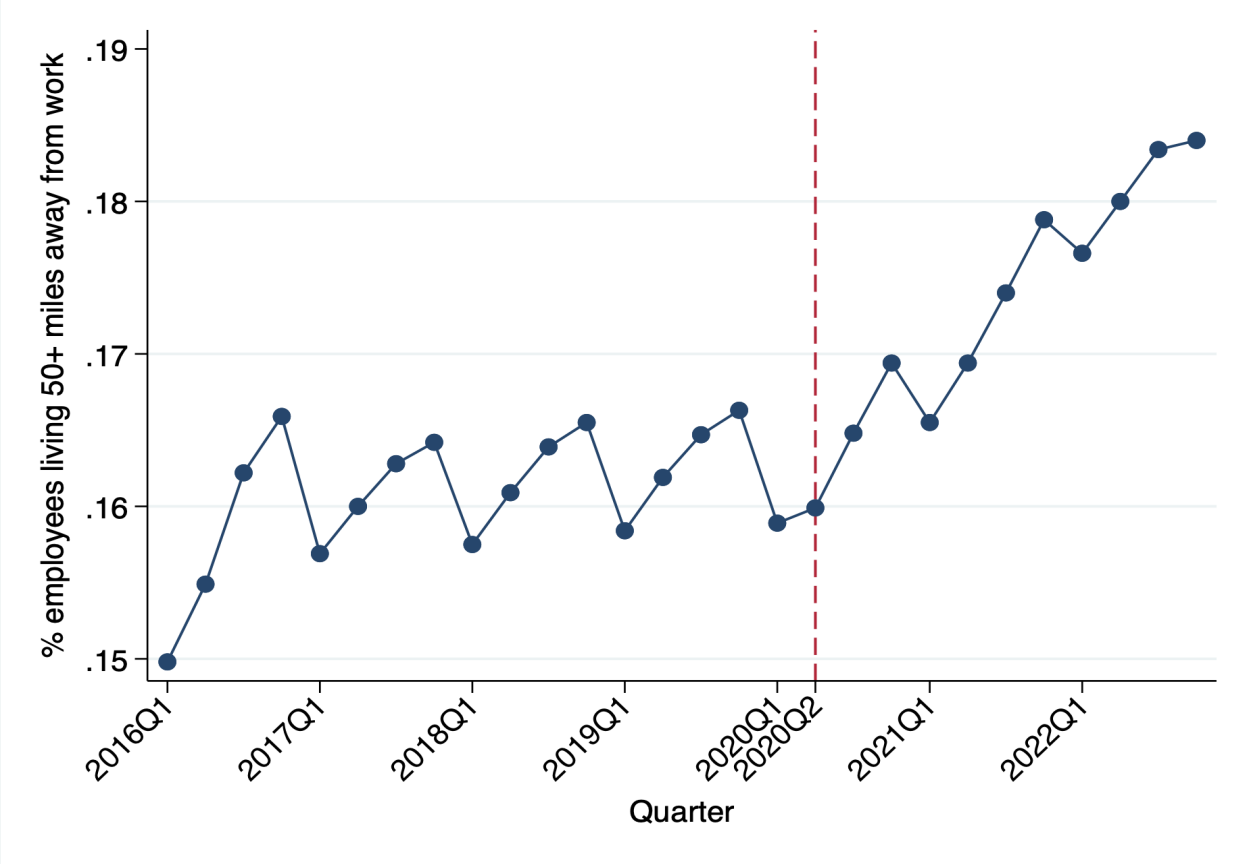


Figure C.3: Percentage of Employees Working 50+ miles from Home, quarterly

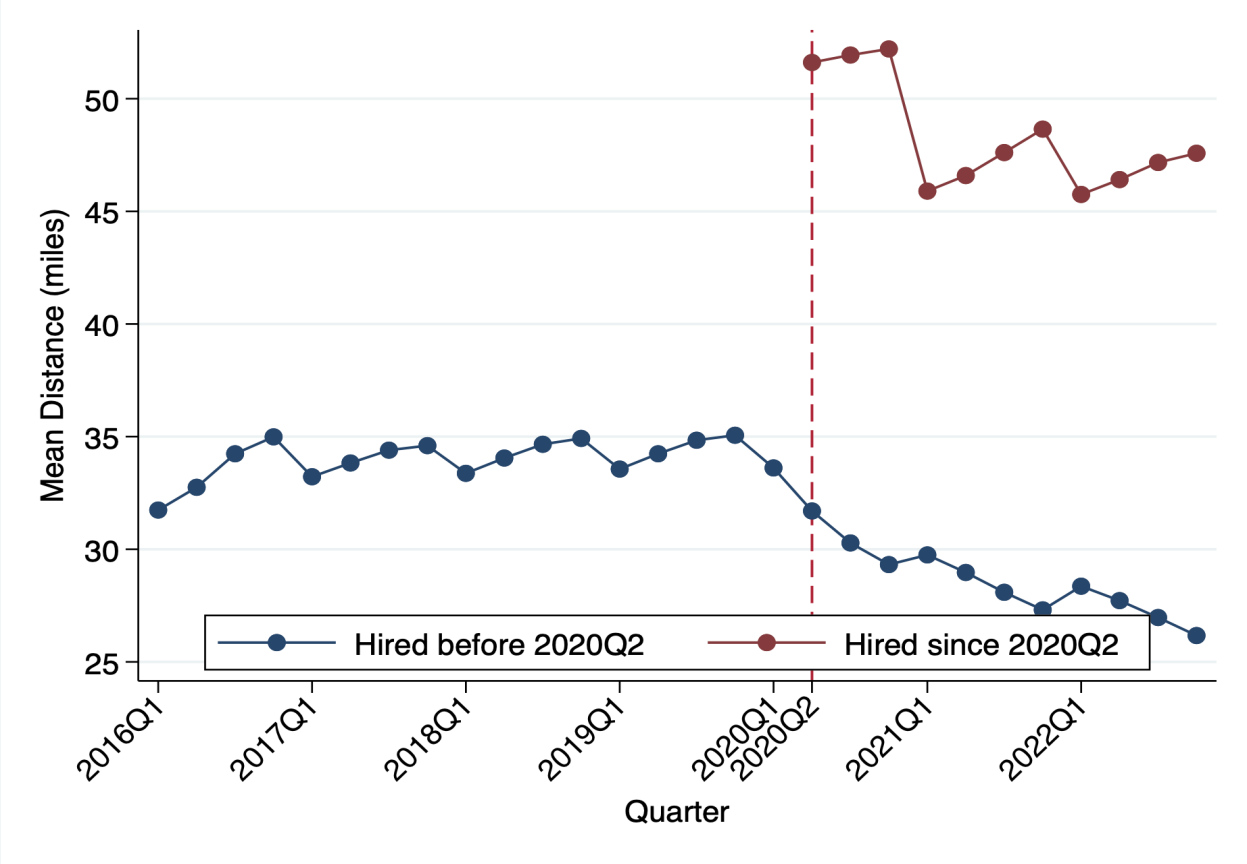


Figure C.4: Average Distance, by hire cohort, quarterly

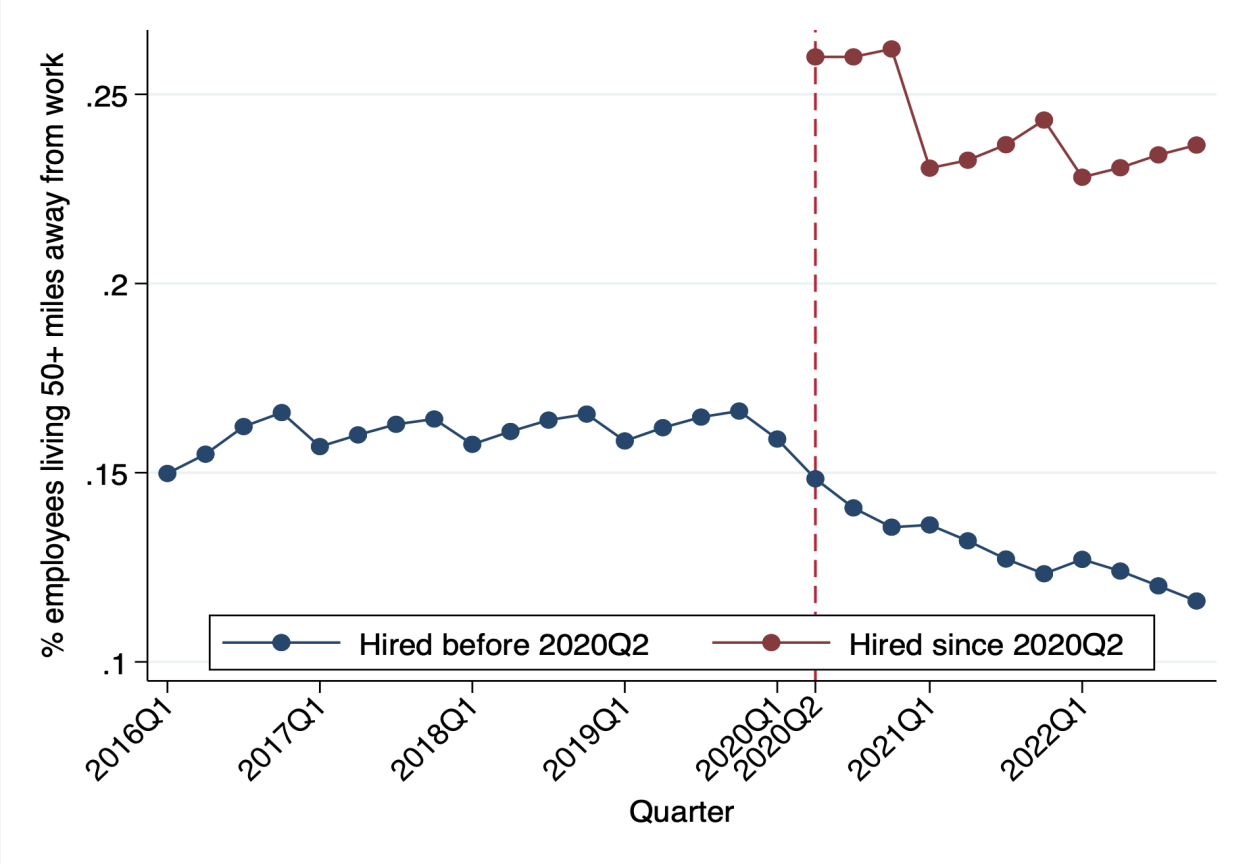


Figure C.5: Percentage of Employees Living 50+ miles from Work, by hire cohort, quarterly

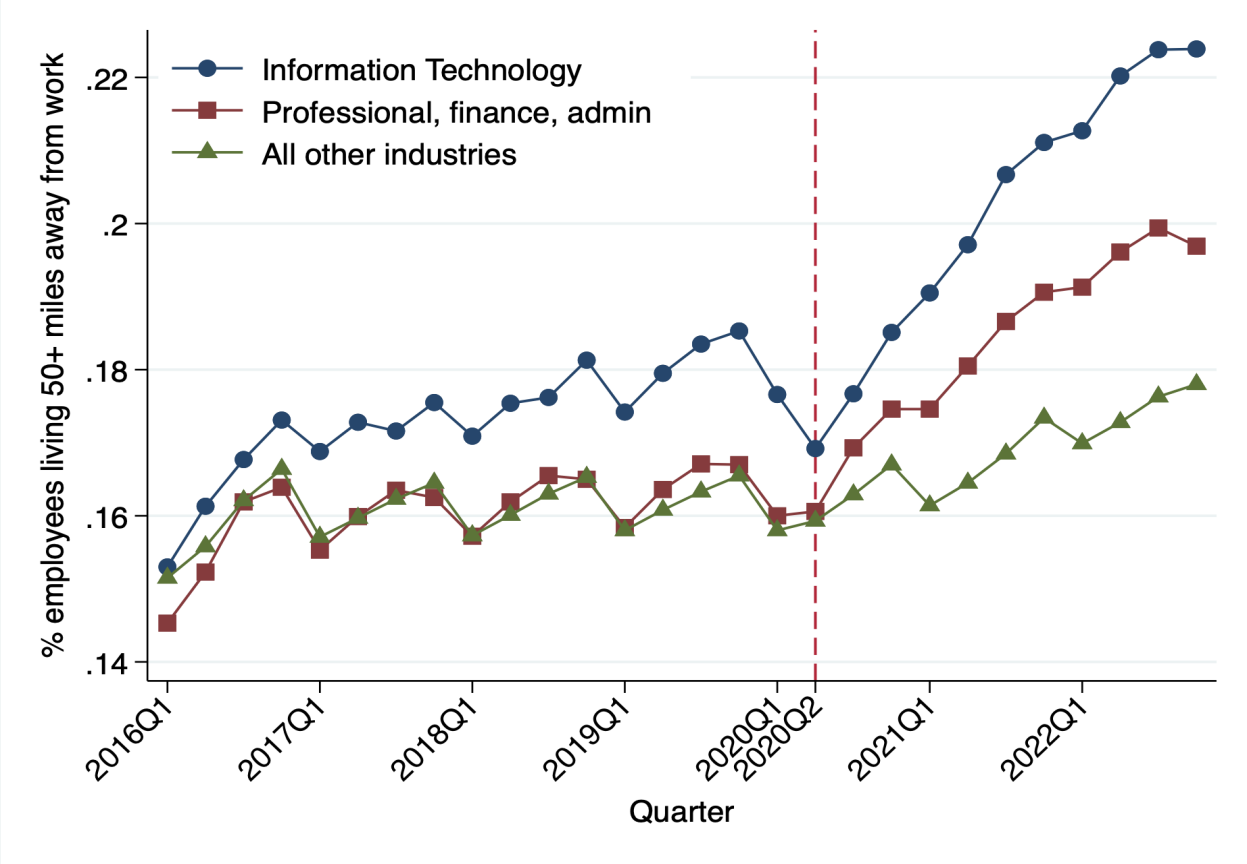


Figure C.6: Percentage of Employees Living 50+ miles from Work, by Industry, quarterly

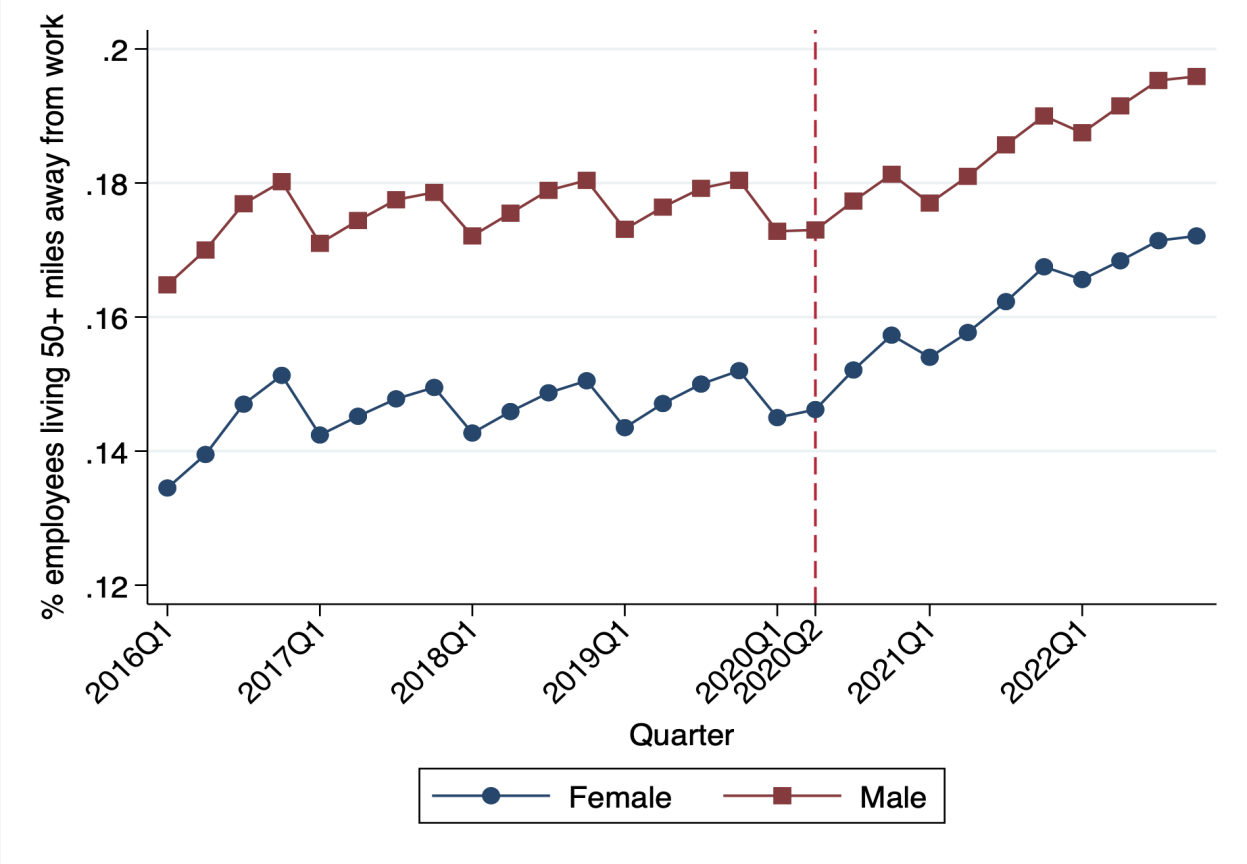


Figure C.7: Percentage of Employees Living 50+ miles from Work, by Gender, quarterly

BIBLIOGRAPHY

K. G. Abraham and L. F. Katz. Cylical unemployment: Sectoral shifts or aggregate disturbances? Journal of Political Economy, 94:3:507–522, 1986.

A. Adams-Prassl, T. Boneva, M. Golin, and C. Rauh. Inequality in the impact of the coronavirus shock: Evidence from real time surveys. Journal of Public Economics, 189, 2020. doi: 10.1016/j.jpubeco.2020.104245.

M. Akan, J. M. Barrero, N. Bloom, T. Bowen, S. Buckman, S. J. Davis, L. Pardue, and L. Wilke. The new geography of labor markets. National Bureau of Economic Research, (33582), 2024.

C. Aksoy, J. M. Barrero, N. Bloom, S. J. Davis, M. Dolls, and P. Zarate. Time savings when working from home. AEA Papers and Proceedings, 113, 2023.

E. Atalay, S. Sotelo, and D. Tannenbaum. The geography of job tasks. Working Paper, 2021.

D. Autor and D. Dorn. The growth of low skill service jobs and the polarization of the u.s. labor market. American Economic Review, 103:5:1553–1597, 2013.

J. Azar, E. Huet-Vaughn, I. Marinescu, B. Taska, and T. von Wachter. Concentration in us labor markets: Evidence from online vacancy data. Labour Economics, 66:101866, 2020.

J. M. Barrero, N. Bloom, and S. J. Davis. Why working from home will stick. NBER Working Paper, (28731), 2021. URL https://www.nber.org/papers/w28731.

P. J. Barwick, Y. Liu, E. Patacchini, and Q. Wu. Information, mobile communication, and referral effects. American Economic Review, 113:5:1170–1207, 2023.

T. Baseler. Hidden income and the perceived returns to migration. American Economic Journal: Applied Economics, 15:4:321–352, 2023.

P. Bayer, S. L. Ross, and G. Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. Journal of Political Economy, 116:6:1150–96, 2008.

G. Becker and K. Murphy. The division of labor, coordination costs, and knowledge. Quarterly Journal of Economics, 107:4:1137–1160, 1992.

O. Blanchard and P. Diamond. The beveridge curve. Brookings Papers on Economic Activity, 1989:1:1–76, 1989.

J. Blumenstock, G. Chi, and X. Tan. Migration and the value of social networks. Review of Economic Studies, forthcoming.

D. Borowczyk-Martins, G. Jolivet, and F. Postel-Vinay. Accounting for endogeneity in matching function estimation. Review of Economic Dynamics, 16:3:440–451, 2013.

K. Burdett, S. Shi, and R. Wright. Pricing and matching with frictions. Journal of Political Economy, 109:5:1060–1085, 2001.

S. Caldwell. Outside options, bargaining, and wages: Evidence from coworker networks. Working Paper, 2019.

W. J. Carrington, E. Detragiache, and T. Vishwanath. Migration with endogenous moving costs. American Economic Review, 86:4:909–930, 1996.

F. Cingano and A. Rosolia. People i know: Job search and social networks. Journal of Labor Economics, 30:2:291–332, 2012.

E. DeFilippis, S. M. Impink, M. Singell, J. T. Polzer, and R. Sadun. Collaborating during coronavirus: The impact of covid-19 on the nature of work. NBER Working Paper, (27612), 2020. URL https://www.nber.org/papers/w27612.

D. Deming and L. B. Kahn. Skill requirements across firms and labor markets: Evidence from job postings for professionals. Journal of Labor Economics, 36:337–369, 2018.

M. W. Elsby, R. Michaels, and D. Ratner. The beveridge curve: A survey. Journal of Economic Literature, 53:3:571–630, 2015.

A. Flaaen, M. Shapiro, and I. Sorkin. Reconsidering the consequences of worker displacements: Firm versus worker perspective. American Economic Journal: Macroeconomics, 11:2:193–227, 2019.

S. Frohnweiler, B. Beber, and C. Ebert. Information frictions, belief updating and internal migration: Evidence from ghana and uganda. IZA Discussion Paper No. 15826, 2022.

C. Gaubert, P. Kline, D. Vergara, and D. Yagan. Trends in u.s. spatial inequality: Concentrating affluence and a democratization of poverty. AEA Papers and Proceedings, 111: 520–525, 2021.

M. Graham, E. McEntarfer, K. McKinney, S. Tibbets, and L. Tucker. Lehd snapshot documentation. Working Papers 22-51, Center for Economic Studies, U.S. Census Bureau, 2022.

A. Gupta, V. Mittal, and S. Van Nieuwerburgh. Work from home and the office real estate apocalypse. Working paper, 2024.

J. Hazell and B. Taska. Downward rigidity in the wage for new hires. Working Paper, 2019.

J. Hellerstein, M. McInerney, and D. Neumark. Neighbors and coworkers: The importance of residential labor market networks. Journal of Labor Economics, 29:4:659–695, 2011.

B. Hershbein and L. B. Kahn. Do recessions accelerate routine-biased technological change? evidence from vacancy postings. American Economic Review, 108:7:1737–1722, 2018.

B. J. Hershbein. The new hires quality index: A wage metric for newly hired workers. W.E. Upjohn Institute for Employment Research, 2017.

Y. Ioannides and L. Loury. Job information networks, neighborhood effects, and inequality. Journal of Economic Literature, 42:4:1056–1093, 2004.

N. Jia, R. Molloy, C. Smith, and A. Wozniak. The economics of internal migration: Advances and policy questions. Journal of Economic Literature, 61:1:144–180, 2023.

P. Kline and E. Moretti. Placed based policies with unemployment. American Economic Review, 103:3:238–243, 2013.

F. Kramarz and O. N. Skans. When strong ties are strong: Networks and youth labour market entry. Review of Economic Studies, 81:3:1164–1200, 2014.

R. Lagos. An alternative approach to search frictions. Journal of Political Economy, 108:5: 851–873, 2000.

B. Lester, D. A. Rivers, and G. Topa. The heterogeneous impact of referrals on labor market outcomes. Working Paper, 2023.

D. M. Lilien. Sectoral shifts and cyclical unemployment. Journal of Political Economy, 90:4: 777–793, 1982.

I. Marinescu and R. Wolthoff. Opening the black box of the matching function: The power of words. Journal of Labor Economics, 38:2:535–568, 2020.

A. Marshall. The Principles of Economics. New York: MacMillan, 1890.

E. Moretti. Local multipliers. American Economic Review: Papers and Proceedings, 100:2: 373–377, 2010.

E. Moretti. The New Geography of Jobs. Boston: Houghton Mifflin Harcourt, 2012.

K. Munshi. Social networks and migration. Annual Review of Economics, 12:503–524, 2020.

R. Nunn, J. Shambaugh, and J. Parsons. The geography of prosperity. In Place-Based Policies for Shared Economic Growth, pages 11–42. 2018.

B. Petrongolo and C. Pissarides. Looking into the black box: A survey of the matching function. Journal of Economic Literature, 39:390–431, 2001.

C. Porcher. Migration with costly information. Working Paper, 2022.

A. Ramani, J. Alcedo, and N. Bloom. How working from home reshapes cities. Proceedings of the National Academy of Sciences, 121(45), 2024.

A. Sahin, J. Song, G. Topa, and G. L. Violante. Mismatch unemployment. American Economic Review, 104:11:3529–3564, 2014.

I. Schmutte. Job referral networks and the determination of earnings in local labor markets. Journal of Labor Economics, 33:1:1–32, 2015.

R. Shimer. Mismatch. American Economic Review, 97:4:1074–1011, 2007.

A. Smith. An Inquiry into the Nature and Causes of the Wealth of Nations. New York: Modern Library, 2000.

B. Stuart and E. Taylor. Migration networks and location decisions: Evidence from us mass migration. American Economic Journal: Applied Economics, 13:3:134–175, 2021.

L. Tian. Division of labor and productivity advantage of cities: Theory and evidence from brazil. Working Paper, 2018.

U.S. Census Bureau. United states commuting at a glance: American community survey 1-year estimates. https://www.census.gov/topics/employment/commuting/guidance/acs-1yr.html, 2024. Last revised September 9, 2024.

M. Wallskog. Entrepreneurial spillovers across coworkers. Working Paper, 2023.

R. Wilson. Moving to economic opportunity: The migration response to the fracking boom. Journal of Human Resources, 57:3:918–55, 2022.

M. Zhu. Beyond the old boys network: Social networks and job finding at community colleges. Journal of Human Resources, forthcoming. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.0921-11888R2. URL http://jhr.uwpress.org/lookup/doi/10.3368/jhr.0921-11888R2.

S. Zimmerman. Elite colleges and upward mobility to top jobs and top incomes. American Economic Review, 109:1:1–47, 2019.