Cities in the Information Age

Are Information-Intensive Firms

Contributing to Urban Inequality?

Siddarth Shankar

Advisor: Sun Kyoung Lee

A Senior Thesis submitted in partial fulfillment of the requirements of the degree of Bachelor of Arts in Economics

April 6, 2022

Department of Economics

Yale University

New Haven, CT

Contents

Acknowledgments **4**

Abstract **5**

1 Introduction **6**

1.1 Technology's Role in the Rise of the City 7

1.2 Definitions 11

1.3 Characteristics of Information-Intensive Firms 13

2 Relevant Literature **18**

2.1 Information-Intensive Industry Growth & Spatial Inequality 18

3 Theoretical Model **22**

3.1 Simple Model 23

3.2 A Model with Skill-Biased Technological Change 24

4 Data **31**

4.1 Description of Empirical Setting 31

4.2 Firm Data 32

4.3 Worker Data 33

5 Results **36**

5.1 Information-Intensive Industry Growth 37

5.2 Wage Effects 40

5.3 Amenity Effects 43

5.4 Housing Effects 46

5.5 Welfare Effects and Evaluation of the Model 50

6 Discussion and Conclusion **51**

References **54**

A Appendix **57**

A.1 Data Cleaning 57

A.2 Figures 59

A.3 Tables 63

Acknowledgments

Thank you to my advisor, Sun Kyoung Lee, for encouraging me to pursue my intellectual passions and for providing advice and support over the past year. Her guidance has been instrumental in producing this project.

Thank you to my parents, Venkat and Sadhna Shankar, for instilling in me the values of hard work, perseverance, and the pursuit of knowledge. For the past 22 years, their constant sacrifice and their regular encouragement have made me into the person and scholar I am today.

I would be remiss if I failed to note that the world has been an unusually troubling and chaotic place over the past two years, with numerous disruptions to the educational milieu of Yale College. The resilience and fortitude I mustered would not be possible without my friends, who made me feel normal in the most abnormal of times. Thank you.

Projects like these can never be perfect: all errors are my own.

Abstract

Information-intensive firms now generate almost 25% of U.S. GDP and are spatially concentrated in cities. Using a heteroskedasticity-robust fixed effects model on panel data from County Business Patterns from 2003-2019, this paper analyzes the effect of information-intensive job growth on wages, amenities, and housing costs for high- and low-skilled workers in metropolitan statistical areas across the United States. Information-intensive job growth is associated with significant increases in the wage premium and housing costs, leading to increased welfare discrepancies and urban inequality. Currently, cities are investing in efforts to attract information-intensive firms. Perhaps economists and policymakers alike should be aware of skill-based welfare inequality associated with information-intensive firms' disproportionate presence in a small number of cities.

1 Introduction

When Amazon, the largest Internet-based company in the world, announced in 2017 that it would be spending $5 billion on building a second headquarters (HQ2) to host 50,000 high-skilled, high-paying jobs (Amazon.com, Inc. (2017)), two-hundred and thirty-eight (238) cities across North America jumped at the opportunity to become the beneficiaries of Big Tech's big money, submitting initial bids in hopes of convincing Amazon to locate in their cities (O'Connell (2017)). By November 2018, Amazon had settled on its much-anticipated final decision. It would split its HQ2 across two locations with 25,000 workers each: Arlington County, Virginia, and Queens, New York (Amazon.com, Inc. (2018)).

However, persistent local opposition to Amazon's construction in Queens led to the cancellation of the HQ2 project (Amazon.com, Inc. (2019)). Many locals worried that the presence of Amazon's large headquarters would lead to gentrification, as new residents with higher wages would increase demand for housing, pushing out residents by raising their rents. Do these concerns actually bear out in real-world evidence?

The location decisions of Amazon and other information-intensive companies raises an important question for the future of urban economics: **Has the presence of information-intensive companies in cities resulted in differential welfare impacts for high-skilled workers versus low-skilled workers?** While not every company is an Amazon-level behemoth, and not every city is New York City, information intensive companies have emerged as a powerful force in the past two decades, increasingly monopolizing the public's attention and the market's dollars, shaping the cities we live in — for better or for worse. This paper attempts to answer that question through a panel data analysis of U.S. metropolitan areas from 2003-2019.

1.1 Technology's Role in the Rise of the City

In the 21st century, the city is the hub of economic activity and knowledge generation. Across the world, rural-to-urban migration in pursuit of greater economic opportunity continues to increase. But cities were not always economically productive or efficient. Technology has played a key role in the rise of the city. Historically, advances in technology have made cities more accessible and inhabitable. Technologies such as as water management, architecture, and commuting improvements increased the sanitation and density of cities, creating space for more laborers to work together in safer conditions. The technology-driven agglomeration of labor within the city resulted in measurable improvements in productivity that have made metropolitan areas wealthier and better-resourced than their rural counterparts.

The dominant pattern of economic growth in the past two centuries has shown a remarkable urban bias (Eckert et al. (2020)). However, the extent to which the urban bias of growth has manifested has differed over time, and has largely depended on the industry characteristics of urban firms. The recent proliferation of information-intensive companies in cities and the process of computerization and digitization is comparable to other urban-biased productivity shocks that have changed the trajectory of American urban development over the past two centuries (Boustan et al. (2018); Eckert et al. (2020)).[1] In urban economics, a consistent pattern has emerged: technological changes result in increased productivity for cities, leading to the establishment of an urban hierarchy, where certain cities become larger and more productive than others.

[1] American economic historians have observed two distinct urban-biased productivity shocks. The first of these shocks occurred when American cities evolved from small trading sites for local agricultural goods into large regional shipping ports, serving as a conduit of goods and services (Boustan et al. (2018)). The second urban-biased productivity shock occurred when cities became the site of not just regional shipping, but industrial production and manufacturing activity (Boustan et al. (2018)). The Industrial Revolution led to the emergence of cities that produced and shipped goods, increasing prosperity for urban residents. This transformation continued well into the late 20th century. For example, Hornbeck and Moretti (2021) find that productivity growth in the U.S. manufacturing industry in the 1980s and 1990s reduced inequality and raised wages for non-college-educated workers.

How exactly cities emerge and remain dominant in the urban hierarchy is relevant to our research question of interest. Davis and Dingel (2019) find that as cities grow larger, they inherently attract a more skilled set of laborers. Due to comparative advantage, as skill-intensive industries such as information-intensive firms agglomerate in specific locations, these locations will eventually become more attractive to and productive for laborers, leading to a greater willingness to pay for the amenities of the location in question. These cities will thus grow the fastest and become larger over time.

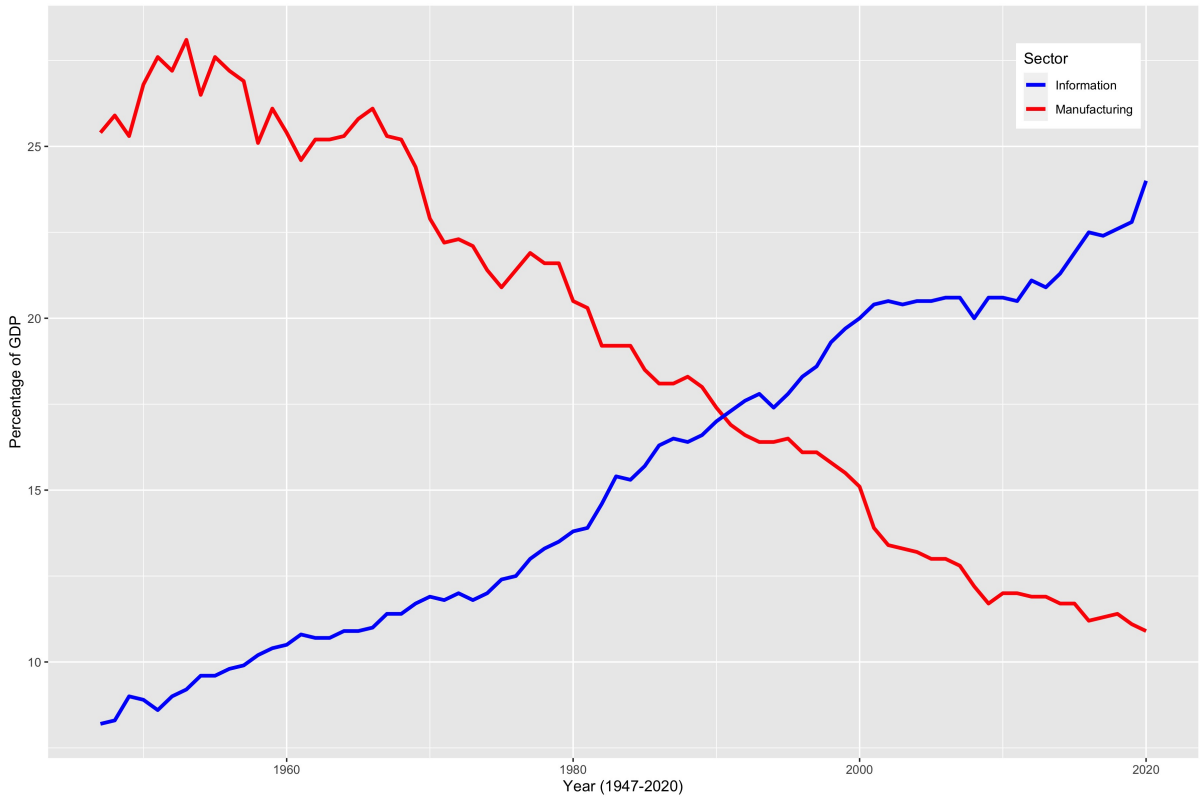
In the United States during the 19th and 20th centuries, technological advances in manufacturing accelerated urban development. By 1900, manufacturing-heavy cities such as St. Louis, Baltimore, Buffalo, Pittsburgh, Cleveland, Milwaukee, and Detroit rose to the top of the urban hierarchy and were among the 15 largest cities in the country, amassing large urban populations and prosperity (Gibson (1998)). But as quickly as industries rise, they can also fall. Manufacturing and industrial production no longer functions as the dominant economic paradigm in the United States. Today, reflecting the decline of the manufacturing sector, none of these cities are among the 15 largest.

Instead, high-skilled, information-intensive jobs have fueled economic growth. These jobs include professions as varied as software and web developers, investment banking analysts, corporate management, and other professional services occupations. This growth has brought about the rise of cities such as San Francisco, San Jose, Los Angeles, and Austin into the 15 largest cities in the country.

The economic literature notes that cities are the beneficiaries of "first nature" and "second nature" causes (Cronon (1992)). While "first nature" causes include geographic features such as topography, waterways, climate, and more, "second nature" causes include agglomeration economies and other supportive human infrastructure. I argue that information is now an important "second nature" force that functions as supportive infrastructure that not only changes the city, but also changes the welfare of people within the city. If information-intensive firm presence indeed enhances welfare for all city residents, forces of path dependence also suggest that cities which invest in and attract information-intensive firms, jobs, and infrastructure are most likely to maintain their status within the urban hierarchy over time (Boustan et al. (2018)).

While historical data is only available from 1947 onwards, the secular declining trend of manufacturing as a share of U.S. gross domestic product (GDP) over time is evident, while a steady, upward rise in information-intensive industries as a share of GDP can be observed, as demonstrated in Figure 1 (U.S. Bureau of Economic Analysis (2022)). A better understanding of how these trends shape workers and cities is crucial to developing better public policy.

Figure 1: Manufacturing and Information-Intensive Sectors as a % of U.S. GDP, 1947-2020



Note: This figure reports the dominance of manufacturing and information-intensive sectors in the U.S. economy over time. Manufacturing statistics include the production of both durable and nondurable goods. Information-intensive statistics comprise of 2022 NAICS Codes 51, 52, 54, and 55 (U.S. Census Bureau (2022b)). GDP is measured in nominal terms. Data is only available since 1947. Source: *U.S. Bureau of Economic Analysis (2022)*.

As the case of Amazon illustrates, the increasing competition between cities to attract information-intensive firms and their high-skilled, highly-paid employees to locate there is a paradigm shift: just as manufacturing firms' location decisions in the 19th and 20th centuries led to the rise of Pittsburgh, Cleveland, and Detroit, information-intensive

firms' location decisions today may dictate the dominant cities of tomorrow.

In the past, information-intensive firms may have chosen the location that maximized their profit through the lowest-cost and most-productive combination of rents, wages, and amenities that a particular city or region has to offer. These firms could be described as *city-takers*, just like a perfectly competitive firm in microeconomic theory is a *price-taker*. Firms would locate in the city as-is, with largely exogenous regional differences in these three categories of rents, wages, and amenities determining their location decisions.

The advent of urban planning and municipal governance, however, has dramatically changed the market for cities, in which buyers (firms) are now directly engaging with sellers (cities) to find the most efficient location for their business. Workers are cut out of the equation, despite being the ones who may benefit or be harmed the most.

The presence of information-intensive companies impacts workers through the channel of local productivity. This change in local productivity affects the wages of all workers of any skill level, or the wage gradient. Furthermore, from the perspective of existing residents, as seen with Amazon and Queens, the presence of such companies may also affect housing affordability and the rent gradient. These factors in conjunction may serve to deepen inequality between high-skilled workers and low-skilled workers.

Just as technological advancements in water management, architecture, and commuting transformed the city of the past, so too do information-intensive industries have the potential to build the cities of the future. But a city cannot exist without workers, and the location decisions of information-intensive companies have the potential to create a new urban hierarchy, with staggering welfare implications for both high-skilled and low-skilled workers. Reckoning with the consequences of the presence of information-intensive firms in cities is an important economic and public policy question that this paper aims to fully explore.

Throughout history, the American city has functioned in many roles: an agricultural market, a shipping port, and a manufacturing hub. With the rise of information-intensive industries, the city is now evolving to its new role as the site of information.

1.2 Definitions

For the purposes of this paper, I define *information-intensive firms* as those firms with 2022 North American Industry Classification System (NAICS) codes 51, 52, 54, and 55 which broadly include occupations that require large amounts of information (U.S. Census Bureau (2022b)). These firms are establishments that, according to their 2-digit NAICS codes, are in the business of:

High-Skilled Industry Definitions by 2022 2-Digit NAICS Codes

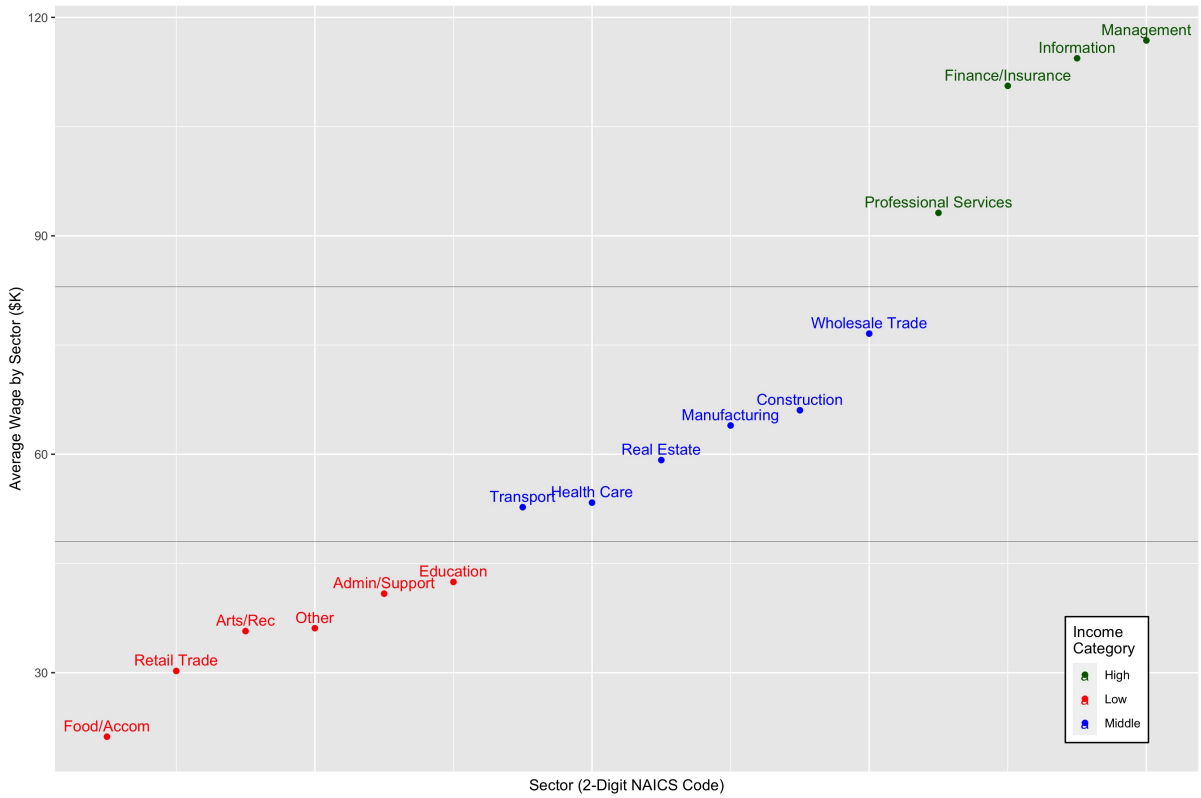
- 51: Information
- 52: Finance and Insurance
- 54: Professional, Scientific, and Technical Services
- 55: Management of Companies and Enterprises

I define such firms as information-intensive because they have either made information into a commodity that can be bought and sold in the market of goods and services or require the rapid transmission and bundling of information in order to sell other assets or make business decisions, i.e. trading a stock, selling a bond, aggregating data, etc.

I also define *high-skilled workers* as any worker that works for an information intensive establishment as defined above. Empirical data supports the idea of a skilled wage premium, and the four industries I define as information-intensive are also the top four industries by average wage in metropolitan statistical areas across the country in industries that employ over one million people, as demonstrated by Figure 2. The four high-skilled industries defined above each employ between 3.1 and 8.1 million people total across all metropolitan statistical areas (U.S. Census Bureau (2022a)). In short, the presence of high average wages in a specific industry implies a high level of skill, and thus forms the basis for the assumption of high-skilled workers among information-intensive firms.

The logical extension of that assumption, as also supported by data from Figure 2, is to then assume that the presence of low average wages in a specific industry implies

Figure 2: Industries Ranked by Average Wage across all Metropolitan Statistical Areas, 2019



Note: This figure reports the average wage of workers in different industries in metropolitan areas across the United States as defined by their 2-digit NAICS Codes. The solid horizontal lines represent the visible terciles of wage heterogeneity. Industry codes are defined by two-digit 2022 North American Industry Classification System (NAICS) Codes (U.S. Census Bureau (2022b)). Only industries that employ over 1 million people across all U.S. MSAs are included (leading to the dropping of 3 industries: (1) Utilities; (2) Mining, Quarrying, and Oil and Gas Extraction; and (3) Agriculture, Forestry, Fishing and Hunting.) Average wage is calculated by dividing annual payroll by the number of employees in each industry for all metropolitan statistical areas across the country. Data from establishments outside of metropolitan statistical areas has been omitted. Source: *U.S. Census Bureau (2022a)*.

a low level of skill. I thus define *low-skilled workers* to be a worker employed in an establishment of any of the following industries as defined by their 2-digit NAICS Code (U.S. Census Bureau (2022b)):

Low-Skilled Industry Definitions by 2022 2-Digit NAICS Codes

- 44: Retail Trade
- 71: Arts, Entertainment, and Recreation
- 72: Accommodation and Food Services
- 81: Other Services (except Public Administration)
- 56: Administrative and Support and Waste Management and Remediation Services
- 61: Educational Services

These low-skilled industries each employ between 2.1 to 12.8 million people (U.S. Census Bureau (2022a)). Note that this assessment of low- and high-skill is not a value judgment of the real skills of different workers, but a reflection of observed patterns of the wage premium of information-intensive or high-skilled workers (as can also be seen in Figure A.2). Figure 2 also displays a clear tercile pattern between the industries that are high-paying and those which pay lower wages, adding an empirical justification for the definition and classification of industries as high-skilled and low-skilled.

1.3 Characteristics of Information-Intensive Firms

Given its historical dominance, the economics literature has largely focused on the manufacturing sector as the primary generator of economic growth in cities. But given the shift away from manufacturing, it is important to understand how economic assumptions made about manufacturing firms differ from the realities of information-intensive firms. The industry characteristics of firms in these two sectors are quite distinct.

Information-intensive firms value locations differently than manufacturing firms in four distinct ways. First, their consumption of land is significantly different: while manufacturing firms might require lots of land to accumulate physical capital such as machinery, the vast majority of information-intensive companies' infrastructure is hosted online; their servers would require significantly less physical space than a manufacturing firm's heavy machinery. Information, unlike manufactured goods, is not a physical or

tangible commodity, and consequently most information-intensive firms do not require significant natural resource inputs.

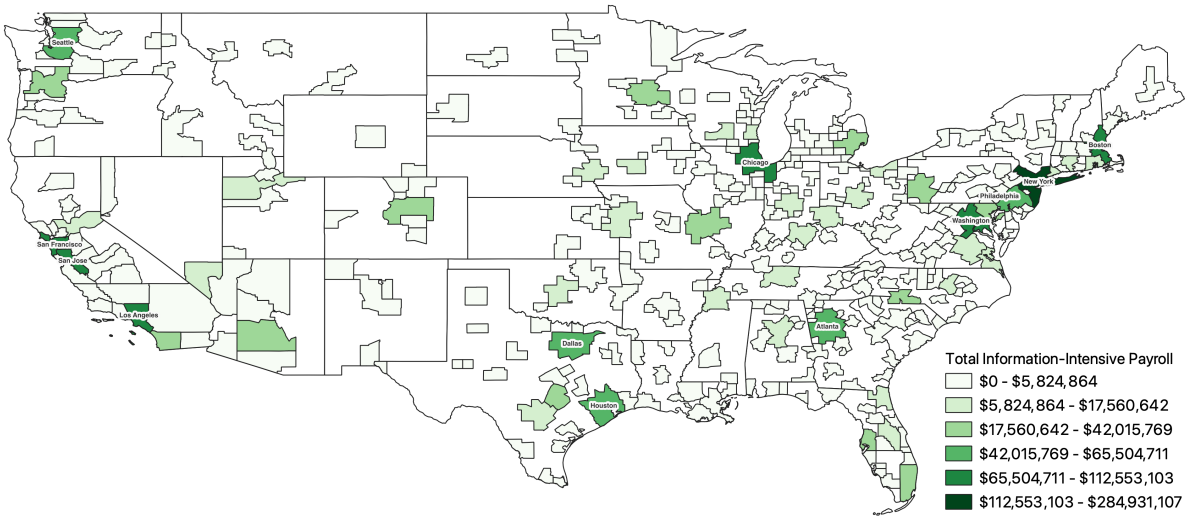
Second, most information-intensive jobs rely on high-skilled, highly-paid laborers with a bachelor's degree or higher, while manufacturing firms do not usually require the same level of educational credentials. This means that information-intensive industries require labor pooling that is optimal to attract high-skilled workers with less search frictions for qualified candidates.

Third, unlike manufacturing firms, information-intensive firms do not need to remain in close proximity to their end consumers due to the far-reaching nature of the Internet. For example, information-intensive firms such as Amazon have already revolutionized the way everyday consumers shop for items, almost completely eliminating city-to-city differences in the variety of items available for purchase (Sinai and Waldfogel (2004)).

Finally, because of the centrality of high-skilled labor in the production process for information-intensive firms, they must be located in close proximity to amenities that high-skilled workers value. Because there is a larger variety of amenities available in cities, the spatial distribution of information-intensive industries is neither random nor uniform, but is very geographically concentrated in cities. Moreover, information-intensive firms are geographically concentrated in *specific* cities: while there are 374 metropolitan statistical areas across the United States, the vast majority of MSAs have negligible presence of information-intensive firms as measured by total payroll (U.S. Census Bureau (2022a)). Figure 3 shows the spatial distribution of these firms.

It is striking that a small subset of these MSAs have a disproportionate presence of information-intensive firms (Table 1). Approximately twenty-five (25) percent of information-intensive firm payroll is concentrated in just three (3) MSAs: New York, Los Angeles, and San Francisco. Moreover, fifty (50) percent of information-intensive firm payroll is concentrated in just ten (10) MSAs, and seventy-five (75) percent of information-intensive firm payroll is concentrated in thirty (30) MSAs.

Figure 3: Total Payroll of Information-Intensive Firms by Metropolitan Statistical Area, 2018



Note: This figure reports the aggregated total payroll of information-intensive firms across the country. Each colored area represents a distinct metropolitan statistical area (MSA), and white areas are not located within any MSA. MSAs with over $50 million in information-intensive payroll are labeled. Geographic shapefiles of each MSA overlaid upon the U.S. states are both derived from U.S. Census Bureau (2010). Legend colors and distinctions are determined through a Jenks natural breaks optimization. Source: *U.S. Census Bureau (2022a)*.

Table 1: Concentration of Information-Intensive Industries in 10 Metropolitan Statistical Areas, 2019

MSA	Inf. Workers	%	% of Inf. Payroll	% of Population
New York	1,913,275	9.2%	13.4%	6.1%
Los Angeles	1,072,494	5.1%	5.3%	4.0%
Chicago	892,150	4.3%	4.5%	2.9%
Washington	841,318	4.0%	4.4%	1.9%
Dallas	704,169	3.4%	3.1%	2.3%
Boston	641,003	3.1%	4.1%	1.5%
San Francisco	628,732	3.0%	5.2%	1.4%
Atlanta	564,414	2.7%	2.7%	1.8%
Philadelphia	541,575	2.6%	2.7%	1.9%
Houston	474,216	2.3%	2.5%	2.1%
Sum of Top 10	8,273,346	39.7%	47.9%	25.9%

Note: This table reports various measures of the spatial concentration of information-intensive firms into a few metropolitan statistical areas (MSAs), including the number of information workers, the percentage of nationwide information workers, percentage of information payroll, and percentage of the total population. Information-specific parameters far exceed the MSAs' percentage of total U.S. population, again reaffirming the spatial concentration of these industries. The U.S. Census Bureau classifies the colloquially-known region of Silicon Valley into two MSAs, one anchored by the city of San Francisco and one by the city of San Jose. Information-intensive statistics comprise of 2022 North American Industry Classification System (NAICS) Codes 51, 52, 54, and 55, which include firms in the information, finance, and insurance industries and firms that manage companies or offer professional, scientific, and technical services (U.S. Census Bureau (2022b)). Source: *U.S. Census Bureau (2022a)*.

Evidence of this spatial concentration can be seen by not just looking at payroll, but also at the number of information-intensive workers by MSA, the percentage of the total information-intensive labor force that lives in each MSA (Figure A.1), and the percentage of total information-intensive payroll in each MSA, which are summarized for convenience in Table 1. Just five (5) MSAs comprise over twenty-five (25) percent of the total information-intensive labor force, an industry which now accounts for roughly 25% of U.S. GDP. Fifteen (15) MSAs comprise over 50% of the information-intensive labor force, and thirty (30) MSAs account for two-thirds (66.7%) of the information-intensive labor force. These patterns reaffirm that a large sector of the American economy is heavily concentrated in a small subset of metropolitan areas.

This spatial concentration of information-intensive firms and workers has broad welfare implications as digital technology continues to transform economic activity. In short, information-intensive firms may be taking advantage of skill-biased technological change to transform the cities where they concentrate, potentially leaving a subset of workers in those cities as well as those in other disfavored cities behind.

There is currently a gap in the literature when it comes to studying the welfare implications of information-intensive firm presence in cities, and this paper contributes to the urban economics literature by attempting to bridge that gap. The crux of the analysis explores how to properly model the welfare implications of information-intensive job growth on high-skilled and low-skilled workers through data on metropolitan statistical areas (MSAs) from the U.S. Census's annual County Business Patterns (CBP) survey.

Through employing a fixed effects model on longitudinal panel data for MSAs across years 2003-2019, I find that the welfare implications of information-intensive firm entrance into a market can be tested against a canonical urban economic model largely based on Rosen (1979) and Roback (1982). Overall, I find that information-intensive firm entrance into an MSA is associated with significant increases to the skilled wage premium, the intrinsic utility value of regions, and housing costs. According to the model, these effects ultimately result in increased welfare discrepancies between workers of high- and low-skill levels, potentially deepening and contributing to further urban inequality.

Section 2 of this paper provides background on information-intensive firms'

historical development in the United States and explores the extant literature in the field. Section 3 postulates a theoretical model to measure the welfare implications of information-intensive job growth for high-skilled and low-skilled workers within a specific MSA. Section 4 explores the data that is used to test the theoretical model developed in Section 3. Section 5 presents the results of the empirical analysis and Section 6 concludes with the policy implications and findings of the paper.

2 Relevant Literature

In this section, I explore the relevant literature regarding my research question of interest and situate this paper in the context of other academic work and empirical studies in the field of urban economics. This review helps reveal trends and gaps in our current understanding of how information-intensive firm presence is associated with spatial inequality. This paper, in particular, emphasizes the idea that studying how worker outcomes differ *within* metropolitan areas is just as important to our understanding of spatial inequality as how worker outcomes differ *between* metropolitan areas.

2.1 Information-Intensive Industry Growth & Spatial Inequality

Macroeconomic theory suggests that the proliferation of information-intensive industries leads to increases in productivity and subsequently *economic* growth, but it remains unclear whether their presence is *welfare-enhancing* for *all* urban residents. A review of the literature finds many studies that examine the idea of total factor productivity growth leading to diverging welfare outcomes among workers in metropolitan areas. Where this paper differs is in its isolation and examination of information industry growth as a driver of this increased productivity growth and studying its potential to generate further spatial inequality among workers of different skill levels.

There is evidence to suggest that information-intensive firm presence may not be positive for all workers. Despite many information-intensive companies' stated aims to break down barriers between people and to reduce communications and transportation

frictions, the presence of these companies, which tend to employ high-skilled, college-educated workers, may actually create more unequal cities (Diamond (2016)). For example, Boustan et al. (2018) find that computerization has been correlated with rising inequality *between* metropolitan and non-metropolitan areas; this paper attempts to discern whether such outcomes might repeat *within* metropolitan areas.

An underlying assumption of urban economic theory from Rosen (1979) and Roback (1982) is the idea that consumer welfare in a location can be calculated as a combination of wages, or the monetary value of employment, and amenities, which consist of the non-monetary value of a specific location. In this conceptual framework, however, wages converge across cities and subsequently, in equilibrium, welfare is equalized across cities. The Rosen (1979) and Roback (1982) canonical urban economic model thus assumes that locations can either have high levels of wages or high levels of amenities, but not both if welfare is to be equalized across locations: in short, workers will be compensated at higher rates to live in areas with a lower level of amenities. Thus, in a model with skill heterogeneity, workers of different skill levels may migrate between cities depending on their preference for higher wages or higher amenities, but ultimately, the welfare equalization assumption will be maintained across skill levels in cities. However, if information-intensive firm presence results in differential productivity advantages and wage growth for certain cities, these welfare equalization assumptions are difficult to uphold. Indeed, empirical data suggests that this pattern no longer holds true.

Diamond (2016) explores panel data from the United States between 1980 and 2000 and finds that over time, more productive cities attracted a greater share of high-skilled workers and saw increases in both wages and amenity values for consumers over time. She concludes that, in order for this to be consistent with the Rosen (1979) and Roback (1982) model, the amenity preferences of workers of different skill levels must have significantly diverged over time, creating a pattern of residential sorting by skill level, with various cities becoming disproportionately high-skilled or low-skilled. This creates spatial inequality *between* cities of high-productivity and high-skill versus those of low-productivity and low-skill, but ultimately leaves unanswered the question of how exactly these changes affect workers *within* these high- and low-productivity areas.

There are a number of reasons why focusing on worker outcomes within metropoli-

tan areas affected by information-intensive firm proliferation is a compelling area of study. The first relates directly to Diamond (2016)'s observations of different revealed preferences for worker locations among college-educated versus non-college-educated workers. The productivity differences brought about by information-intensive firms may be facilitating differential migration patterns by workers of different skill levels between locations (Hornbeck and Moretti (2021)). Cities that attract information-intensive firms may inherently attract higher "quality" workers, i.e. workers who also have a college education, pushing non-college-educated workers out of cities toward other employment opportunities (Boustan et al. (2018)). Moretti (2012) coins the term of the "Great Divergence" to describe the different location choices between college-educated/high-skilled and non-college educated/low-skilled workers.

Empirically, this concept can be observed in the information wage premium: in every single MSA within the United States, the average information-intensive worker wage is higher than the average non-information-intensive worker wage. However, as Figure A.2 displays, the information wage premium is significantly higher in certain areas than in others, which may be facilitating differential in-migration by high-skilled workers and out-migration by low-skilled workers.

A second compelling rationale is related to the unique industry characteristics of information-heavy industries themselves. As Moretti (2012) finds, for such firms, agglomeration forces are much stronger than dispersion forces due to labor pooling effects and knowledge spillovers. Because high-skilled industries require similar skill sets from their laborers, they highly benefit from agglomeration-influenced productivity spillovers (Moretti (2012)). The easy exchange of ideas is particularly powerful for these firms. For information-intensive firms, a prime example is Sand Hill Road in Palo Alto, California, the heart of Silicon Valley. This 5.6 mile stretch of road in close proximity to the San Jose and San Francisco MSAs is home to over 48 venture capital firms who all compete to provide funding to the most promising start-up companies based in the information-intensive space. Information technology firms that wish to acquire funding from these venture capitalists often seek to locate in close proximity to the firms.

This pattern of labor agglomeration can be seen empirically as well in Figure A.3. In most MSAs in the United States, the ratio of information-intensive workers to

non-information workers is far less than 1, but the MSAs with the highest ratios include well-known technology hubs such as Seattle, San Francisco, and San Jose, as well as other educational capitals, such as Boulder and Boston. This data is a clear demonstration that labor pooling — being home to a large number of high-skilled workers – and subsequent knowledge spillovers are especially important for information-intensive industries. Davis and Dingel (2019) explore this idea of a spatial knowledge economy in cities, concluding that idea generation is explanatory of why certain cities have higher ratios of high-skilled laborers to low-skilled laborers.

Under the Davis and Dingel (2019) theory, information-intensive firm presence also magnifies the productivity of other industries, a positive spillover that makes this information-driven urban-biased productivity shock a salient area of study. Thus, a major reason why so many cities bid for Amazon's HQ2 is the spillover effect of these information-intensive firms: landing a big whale like Amazon is thought to benefit other firms due to the proximity and density of high-skilled workers in the region.

However, if the benefits of increased productivity growth are disproportionately concentrated among high-skilled workers, it may be the case that information-intensive firm presence leads to rising inequality. Indeed, findings by Fleck et al. (2011) find that despite robust information-driven productivity growth, median wages in the U.S. are depressed — a smaller share of national income is going toward labor compensation and more is going to investments in capital.

Another way in which information-intensive firm presence may lead to unequal outcomes between high-skilled, technologically literate workers versus low-skilled, technologically illiterate workers, is in driving up the cost of housing, leaving certain cities as enclaves for either the wealthy or the poor. Indeed, findings by Hsieh and Moretti (2019) conclude that productivity differences across U.S. cities, given cities' limited housing supply, has resulted in the spatial misallocation of labor – suggesting that total U.S. welfare is not maximized due to the prevalence of many cities that have either very high or very low skilled labor ratios.

These housing patterns can be observed in Figure A.4, as housing costs are highest among many of the information-intensive regions highlighted in Figure A.3. In this

way, information-intensive firms may be directly contributing to the "Great Divergence" — the exact same phenomenon that many residents of Queens were afraid of during Amazon's HQ2 process.

Overall, the literature suggests that the urban bias of information firms is especially important to study, because unlike most industries, which have experienced rapid de-densification and suburbanization, information-based industries have not decentralized to this extent (Boustan et al. (2018)). As mentioned above, Davis and Dingel (2019) find that the benefits of agglomeration far exceed the costs for information-intensive firms in a way that is untrue for other industries and sectors. It is then reasonable to expect that information-intensive firms will continue to grow their urban presence over time, underscoring the importance of a deeper understanding of the welfare implications of this pattern on city residents.

Given these unanswered questions and the continued rise of information-intensive industries since Diamond (2016)'s data concluded in 2000, I explore panel data from 2003-2019 to see whether I can ascertain new patterns of spatial inequality or determine whether these observed patterns may have continued, reversed, or accelerated over time. In the next section, however, I briefly explain how economic theory attempts to reconcile these questions and our research question of interest.

3 Theoretical Model

In this section, I elaborate upon a canonical urban economic model from Rosen (1979) and Roback (1982). I then describe adjustments to this simple model that better fit our research question of interest. The amended model introduces the idea of skill-biased technological change, in which the proliferation of information promotes and favors the substitution of low-skilled labor for high-skilled labor in firms' production processes. The model determines that welfare discrepancies between low- and high-skilled workers arise from differences in the skilled wage premium, the ratio of high to low-skilled workers (labor ratio), and differential amenity preferences between workers of differing skill levels.

3.1 Simple Model

A canonical urban economic model developed by Rosen (1979) & Roback (1982) provides a baseline to help explore our question of interest. The model assumes that a single homogeneous good is freely traded across locations in space S with no trade costs. This fits well with the idea of information as a good that, in the age of digital media, is freely traded across space with no trade costs.

Workers

The welfare W of a worker in location j is represented by $W_j = C_j u_j$ subject to budget constraint $p_j C_j = w_j$, where C_j represents consumption in location j, u_j represents the amenity or intrinsic utility value of residing in location j, p_j reflects the price in location j, and w_j reflects the wage in location j. Because there is only one homogeneous good that is freely traded in this model, we can set p_j as our numeraire such that $p_j = p = 1$. Since p_j is normalized to 1, re-arranging the budget constraint yields $C_j = w_j$ where consumption in location j is equal to the wage in location j. Thus, in this model, all worker earnings are spent on consumption of the homogeneous good.

Firms

Firms are perfectly competitive with price equal to marginal cost of production, or $p_j = w_j/A_j$, where A_j is the productivity of the location j. Because of the properties of the numeraire, where $p_j = p = 1$, we then have that $w_j = A_j$ such that the wage in location j is equal to the productivity of location j. The firm's production function is $Y_j = A_j L_j$ which can be expressed as the *market-clearing condition* $Y_j = w_j L_j$.

Adding Labor, Agglomeration, and Dispersion Forces

Productivity A_j and amenity u_j are functions of intrinsic productivity $\bar{A}_j$, intrinsic amenity $\bar{u}_j$, labor L_j, and forces of agglomeration α and dispersion β. Specifically $A_j = \bar{A}_j L_j^{\alpha}$ and $u_j = \bar{u}_j L_j^{-\beta}$. Because $\frac{\partial A_j}{\partial L_j} > 0$ and $\frac{\partial u_j}{\partial L_j} < 0$, as population increases, the strength of agglomeration forces α will increase productivity, but the strength of dispersion forces β will decrease amenity.

From combining equations, we know that $W_j = A_j u_j$ and subsequently $W_j = \bar{A}_j \bar{u}_j L_j^{\alpha-\beta}$. Welfare W of a worker in location j depends on both the intrinsic amenity and productivity of location j, the number of workers L_j, and the strength of the agglomeration and dispersion forces α and β. We can then update the production function as $Y_j = \bar{A}_j L_j^{1+\alpha}$. Thus, output Y_j is dependent only on the intrinsic productivity $\bar{A}_j$, number of laborers L_j, and the strength of agglomeration forces α.

Equilibrium

There are two equilibrium conditions. The first is the condition of *welfare equalization*, in which welfare across locations must be equalized such that $W_j = \bar{W}$. The second condition is the *aggregate labor constraint*, in which the number of workers across all locations must equal the total population such that $\sum_j L_j = \bar{L}$.

Using these two equilibrium conditions, the labor supply can be computed as $L_j = \bar{W}^{\frac{1}{\alpha-\beta}} \left(\bar{A}_j \bar{u}_j\right)^{\frac{1}{\beta-\alpha}}$. Summing across locations yields $\sum_j L_j = \bar{L} = \bar{W}^{\frac{1}{\alpha-\beta}} \sum_j \left(\bar{A}_j \bar{u}_j\right)^{\frac{1}{\beta-\alpha}}$, which can be rearranged to form $\bar{W} = \frac{\left[\sum_j \left(\bar{A}_j \bar{u}_j\right)^{\frac{1}{\beta-\alpha}}\right]^{\beta-\alpha}}{\bar{L}^{\beta-\alpha}}$.

3.2 A Model with Skill-Biased Technological Change

While the simple Rosen (1979) & Roback (1982) model describes many observed patterns of urban economics, appropriately capturing that locations with higher intrinsic productivity $\bar{A}_j$ and higher intrinsic amenity $\bar{u}_j$ will have a larger population L_j as workers will migrate to these regions accordingly, it is not sufficient for exploring our question of interest.

The model as it stands assumes all workers have homogeneous skill sets, which is untrue in the real world. The equilibrium condition of welfare equalization across locations and workers is also not reflective of reality when observed data shows that there is a large wage premium for workers who are higher skilled and have a bachelor's degree or higher.

A more realistic assumption would be welfare equalization across skill levels. Amending the Rosen (1979) & Roback (1982) model to reflect this concept of *skill-biased*

technological change that information-intensive firms bring to the city is crucial in order to evaluate our question of interest.

Unlike the simple model, the model with skill-biased technological change based off work by Giannone (2011) and Allen and Arkolakis (2014) no longer assumes that a single homogeneous good is freely traded across locations in space S. It incorporates multiple goods with different prices.

The model assumes no iceberg trade costs τ_{ij} or migration frictions κ_{ij} between two locations i and j. While this is a strong assumption, the proliferation of information technology suggests these costs have declined over time, in many cases becoming negligible. As a result, the elimination of these frictions should not dramatically skew our findings based on this model. An explanation of this model accordingly follows.

Workers

The model is predicated on the existence of two distinct skill levels among workers: high-skill level H and low-skill level L. In the context of our research question, a worker with high-skill level H is one that is employed at an information-intensive firm, while a worker with low-skill level L is one that is employed in one of the four industries defined in Section 1.2.

Although the separation of skill level into two discrete categories H and L is by no means entirely reflective of the continuum of skills that workers possess, it simplifies the model to make it accessible.

The welfare W of a worker of skill level M in location i where $H, L \in M$ is represented by

$$W_i^M = C_i^M u_i^M \tag{1}$$

subject to budget constraint

$$P_i C_i^M = w_i^M \tag{2}$$

where C_i^M represents the consumption of workers with skill M in location i, u_i^M represents the intrinsic utility value of residing in location i for a worker of skill M, P_i reflects the price index of all goods in location i, and w_i^M reflects the wage for a worker with skill M in location i.

Re-arranging the budget constraint yields

$$C_i{}^M = \frac{w_i{}^M}{P_i} \tag{3}$$

Thus, in this model, consumption for each level of skilled worker is equal to their wage divided by the price index, otherwise known as the *real wage*.

Firms

Firms are still perfectly competitive with price equal to marginal cost of production. However, accounting for different skill levels M such that $H, L \in M$, we have $P_i = w_i{}^M / A_i{}^M$ where $A_i{}^M$ is the productivity of the location i for workers of skill level M and $w_i{}^M$ is the nominal wage of workers of skill level M. As a result of this, we then have that

$$\frac{w_i{}^M}{P_i} = A_i{}^M \tag{4}$$

such that the real wage in location i is equal to the productivity of location i for skill level M. Combining this with the intuition of Equations (1)-(3), we receive the result

$$W_i{}^M = \frac{w_i{}^M}{P_i} u_i{}^M \tag{5}$$

Unlike the simple model, when we had undifferentiated labor, we must also incorporate an elasticity of substitution between high-skilled and low-skilled labor, ρ. We also want to incorporate a generalized productivity parameter that is not location or skill-specific, $\bar{A}$. For the purposes of our research question, $\bar{A}$ may reflect a total factor productivity metric. With these additions to the model, the firm's production function is

$$Y_i = \bar{A} \left[\sum_{H,L \in M} \left(A_i{}^M L_i{}^M \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \tag{6}$$

which as a result of Equation (4) can be expressed as the *market-clearing condition*

$$Y_i = \bar{A} \left[\sum_{H,L \in M} \left(\frac{w_i{}^M}{P_i} L_i{}^M \right)^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \tag{7}$$

Adding Labor, Agglomeration, and Dispersion Forces

Productivity $A_i{}^M$ and amenity $u_i{}^M$ are functions of intrinsic productivity $\bar{A}_i^M$, intrinsic amenity $\bar{u}_i^M$, labor $L_i{}^M$, and forces of agglomeration α and dispersion β.

Specifically

$$A_i{}^M = \bar{A}_i^M L_i{}^{M\alpha} \tag{8}$$

$$u_i{}^M = \bar{u}_i^M L_i{}^{M-\beta} \tag{9}$$

From Equations (8) and (9), we see that $\frac{\partial A_i{}^M}{\partial L_i{}^M} > 0$ and $\frac{\partial u_i{}^M}{\partial L_i{}^M} < 0$. In other words, as population increases, the strength of agglomeration forces α will increase productivity, but the strength of dispersion forces β will decrease amenity.

From Equations (4) and (5), we know that $W_i{}^M = A_i{}^M u_i{}^M$. Combining this result with Equations (8) and (9) yields

$$W_i{}^M = \bar{A}_i^M \bar{u}_i^M L_i{}^{M\alpha-\beta} \tag{10}$$

Welfare W of a worker of skill M in location i such that $H, L \in M$ depends on both the intrinsic amenity and productivity of location i for a worker of skill M, the number of workers of skill M in location i, $L_i{}^M$, and the strength of the agglomeration and dispersion forces α and β.

From Combining Equations (6) and (8), we can also update the firm's production function as

$$Y_i = \bar{A}\left[\sum_{H,L\in M}\left(\bar{A}_i^M L_i{}^{M1+\alpha}\right)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} \tag{11}$$

Thus, output Y_i is dependent on location- and skill-agnostic productivity $\bar{A}$, skill-based intrinsic productivity $\bar{A}_i^M$, the number of laborers of each skill type $L_i{}^M$, the strength of agglomeration forces α, and the labor elasticity of substitution ρ. Notably, output is not affected by dispersion forces β, but welfare is. This speaks to the strength of agglomeration economies in the face of technological change as well as the disamenity brought to workers when there is too much agglomeration.

Solving the Firm's Problem

With two different types of skilled labor L and H, the firm must decide how to allocate labor in a cost-minimizing way.

The firm's problem can be expressed as

$$\min_{L_i{}^M} \sum_{H,L\in M} w_i{}^M L_i{}^M \tag{12}$$

subject to constraint

$$\left[\sum_{H,L\in M}\left(\bar{A}_i^M {L_i}^{M^{1+\alpha}}\right)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}} = 1 \tag{13}$$

To solve the problem, we can re-write the problem as a Lagrangian and take first-order conditions (FOCs) accordingly.

$$\mathcal{L} = \min_{{L_i}^M} \sum_{H,L\in M} {w_i}^M {L_i}^M + \lambda\left(1 - \left[\sum_{H,L\in M}\left(\bar{A}_i^M {L_i}^{M^{1+\alpha}}\right)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}\right) \tag{14}$$

$$\frac{\partial\mathcal{L}}{\partial {L_i}^H} = {w_i}^H - \lambda\left(\bar{A}_i^H\right)^{\frac{\rho-1}{\rho}}\left({L_i}^H\right)^{\frac{-(1+\alpha)}{\rho}} = 0$$

$$\frac{\partial\mathcal{L}}{\partial {L_i}^L} = {w_i}^L - \lambda\left(\bar{A}_i^L\right)^{\frac{\rho-1}{\rho}}\left({L_i}^L\right)^{\frac{-(1+\alpha)}{\rho}} = 0$$

From Equation (14) and our FOCs, we can derive a calculation for the wages ${w_i}^H$ and ${w_i}^L$ accordingly.

$${w_i}^H = \lambda\left(\bar{A}_i^H\right)^{\frac{\rho-1}{\rho}}\left({L_i}^H\right)^{\frac{-(1+\alpha)}{\rho}} \tag{15}$$

$${w_i}^L = \lambda\left(\bar{A}_i^L\right)^{\frac{\rho-1}{\rho}}\left({L_i}^L\right)^{\frac{-(1+\alpha)}{\rho}} \tag{16}$$

and subsequently can also derive the wage premium for high-skilled workers.

$$\frac{{w_i}^H}{{w_i}^L} = \left(\frac{\bar{A}_i^H}{\bar{A}_i^L}\right)^{\frac{\rho-1}{\rho}}\left(\frac{{L_i}^H}{{L_i}^L}\right)^{\frac{-(1+\alpha)}{\rho}} \tag{17}$$

Combining Equations (4), (5), (8), (10) and (17) allows us to calculate the welfare differential between high and low-skilled workers, as well.

$$\frac{{W_i}^H}{{W_i}^L} = \frac{{w_i}^H}{{w_i}^L}\left(\frac{{L_i}^H}{{L_i}^L}\right)^{-\beta}\left(\frac{\bar{u}_i^H}{\bar{u}_i^L}\right) \tag{18}$$

$$\frac{{W_i}^H}{{W_i}^L} = \left(\frac{\bar{A}_i^H}{\bar{A}_i^L}\right)^{\frac{\rho-1}{\rho}}\left(\frac{{L_i}^H}{{L_i}^L}\right)^{\frac{-(1+\alpha)}{\rho}-\beta}\left(\frac{\bar{u}_i^H}{\bar{u}_i^L}\right) \tag{19}$$

Equilibrium

As before, there are two equilibrium conditions.

The first is the condition of *welfare equalization.* However, in this case, it is not that welfare across locations is equalized; rather, welfare across skill levels is equalized such that for skill level M where $H, L \in M$, ${W_i}^M = \bar{W}^M$.

The second condition is the *aggregate labor constraint*, in which the number of workers across all locations and skill levels must equal the total population. For skill level M where $H, L \in M$, $\sum_i {L_i}^M = \bar{L}^M$.

Using these two equilibrium conditions, Equation (10) can be rearranged to compute the labor supply for each skill level M, ${L_i}^M$ as

$$ {L_i}^M = \bar{W}^{M\frac{1}{\alpha-\beta}} \left(\bar{A}_i^M \bar{u}_i^M\right)^{\frac{1}{\beta-\alpha}} \tag{20} $$

Then, summing across locations, we compute

$$ \sum_i {L_i}^M = \bar{L}^M = \bar{W}^{M\frac{1}{\alpha-\beta}} \sum_i \left(\bar{A}_i^M \bar{u}_i^M\right)^{\frac{1}{\beta-\alpha}} \tag{21} $$

which is rearranged to form

$$ \bar{W}^M = \frac{\left[\sum_i \left(\bar{A}_i^M \bar{u}_i^M\right)^{\frac{1}{\beta-\alpha}}\right]^{\beta-\alpha}}{\bar{L}^{M\beta-\alpha}} \tag{22} $$

Evaluation of the Model

The amended Rosen (1979) & Roback (1982) model with skill-biased technological change as developed by Giannone (2011) and Allen and Arkolakis (2014) is an appropriate model for our research question of interest. As before, it appropriately captures, from Equation (20), that locations with higher intrinsic productivity $\bar{A}_i^M$ and higher intrinsic amenity $\bar{u}_i^M$ will have a larger population of laborers of skill level M, ${L_i}^M$, thereby capturing the differential migration patterns that have been observed in the data, including from Moretti (2012) and Diamond (2016). It not only incorporates the opposing effects of agglomeration α and dispersion β that occur as labor supply ${L_i}^M$ increases, but it also accounts for the elasticity of labor substitution ρ, which has a clear effect on how firms value, hire, and pay workers. The model, through price index P_i accounts for a variety of goods, and not just a single, homogeneous good, thus incorporating the concept of real wage into calculations of welfare and productivity. The equilibrium condition of welfare equalization across skill levels adequately reflects the large wage premium for workers who are higher-skilled and have a bachelor's degree or higher.

Equation (22) also captures the intuition that welfare equalized across skill levels, $\bar{W}^M$, is a function of both the total labor supply for that specific skill level, $\bar{L}^M$, the

sum of intrinsic productivity $\bar{A}^M$ and intrinsic amenity $\bar{u}^M$ across locations $i \in S$ and within skill level M, and agglomeration and dispersion forces α and β.

Finally, Equations (17) and (19) imply that for $\rho > 1$, an event (such as the entrance of an information-intensive firm or adoption of information-intensive technology) that improves $\bar{A}_i^H$ relative to $\bar{A}_i^L$ will exacerbate the skilled wage premium and the skilled welfare differential, thus increasing inequality through the mechanism of skill-biased technological change that benefits high-skilled workers at the expense of low-skilled workers.

This model is certainly a better fit for application to our specific research question. The entrance of information-intensive firms into a location i would directly increase ${L_i}^H$ through hiring of more workers of skill level H. information-intensive firm entrance will also increase intrinsic productivity of high-skilled workers $\bar{A}_i^H$ through knowledge spillovers. Firm entrance may even increase the strength of agglomeration forces α through labor pooling effects.

As mentioned earlier, information-intensive firm entrance has indeterminate effects on the intrinsic utility parameter (or, in our specification, housing costs) $\bar{u}_i^M$ where $H, L \in M$. It could be the case that information-intensive firm entrance causing an increase in ${L_i}^H$ also drives an increase in $\bar{u}_i^H$ due to the community investing in amenities that high-skilled workers prefer and inherently drive up the prices of housing, in which case the entrance of information-intensive firms may positively affect high-skilled workers. However, if too many high-skilled workers move to location i, any increase in $\bar{u}_i^H$ may be negated by the strength of the dispersion forces β.

The story is different for low-skilled workers. The increase in ${L_i}^H$ caused by information-intensive firm entrance decreases the proportion of workers of low-skill level, ${L_i}^L$. Information-intensive technology adoption may exogenously increase the elasticity of labor substitution ρ by making it likelier for firms to demand workers of skill H rather than of skill L. This may inherently disadvantage low-skilled workers who are less likely to be technologically proficient since they are more easily substituted in firms' production processes.

While the model suggests that information-intensive firm entrance will decrease

welfare for low-skilled workers through the mechanism of increased elasticity of substitution ρ, firm entrance could conceivably increase or decrease the intrinsic level of productivity $\bar{A}_i{}^L$ or amenity $\bar{u}_i{}^L$. Without an understanding of the direction and magnitude of the changes that information-intensive firm entrance makes to these parameters, it is impossible to make any conclusive determinations on how information-intensive firm entrance might impact the welfare differential $\frac{W_i{}^H}{W_i{}^L}$. This is where real world data, applied to our model, comes in.

4 Data

In this section, I explain the empirical setting of our research question and analyze the sources of data on the firm level and the worker level that are necessary in evaluating our question of interest when testing it against our theoretical model specification. By combining longitudinal panel data from the U.S. Census Bureau on metropolitan statistical area-level business patterns with Zillow's commercial home value data and self-generated data on the amenity value of metropolitan statistical areas across the country and over time, I ultimately generate a novel data set of observations of all 374 metropolitan statistical areas of the United States across the period 2003-2019. Further details about data collection, cleaning, and processing that are not discussed below are available in the Appendix.

4.1 Description of Empirical Setting

The policy setting of the empirical analysis is the expansion of information-intensive firms in metropolitan statistical areas (MSAs) in the United States between the years 2003 and 2019. The year 2019 is the latest year that data is available, and also ensures that our data is not skewed by any statistical anomalies associated with data collection during the COVID-19 pandemic that began in early 2020. The U.S. Census Bureau defines an MSA to be a region of one or more counties with at least one nuclear city of greater than 50,000 inhabitants and a total population of over 100,000 inhabitants (U.S. Census Bureau (2022a)). Multiple neighboring counties may be classified into the

same MSA around population nuclei if they fit certain criteria, including the commuting patterns, population density, and/or the percentage of urban residents.

MSAs are the appropriate level of aggregation for the purposes of our research question due to the arbitrary nature of municipal boundaries within different U.S. metropolitan areas. Limiting our analysis to what is defined as the official boundaries of a single city or county, rather than the broader metropolitan statistical area, risks influencing our results depending on how certain cities' municipal boundaries are drawn.

Another important reason for using MSAs as our unit of measure is because they spatially and geographically capture the true extent of urban agglomeration, transcending state and county borders. Many major metropolitan areas cross into two or three states, so county and state-level data cannot accurately capture the level of detail that MSA-level data can in these instances.

Additionally, unlike municipal boundaries which can be subject to changes due to city annexation or neighborhood secession efforts, the boundaries of MSAs within the U.S. stay relatively constant over time. The U.S. Census Bureau may make minor revisions to MSA boundaries every ten years depending on changes in the population, population density, or commuting patterns of a specific MSA. Each MSA is assigned a distinct Federal Information Processing Series (FIPS) code, making it easy to both track each MSA over time and spatially map changes in various variables of interest over time.

Moreover, firm- and worker-specific data is often aggregated at the MSA level, which makes intertemporal comparisons across MSAs between variables of interest in our model more feasible. The remainder of this section will discuss the firm- and worker-specific sources of data that are aggregated at the MSA level.

4.2 Firm Data

I look broadly across the metropolitan statistical areas (MSAs) of the United States for the period 2003-2019 to analyze firm-level trends between information-intensive industries and non-information-intensive industries. My main source of data for this is the U.S. Census Bureau's annual County Business Patterns (CBP) data, which contains

details about the number of establishments, total employment, and annual payroll by each different industry type, as defined by the North American Industrial Classification System (NAICS) Codes (U.S. Census Bureau (2022a); U.S. Census Bureau (2022b)). CBP data thus helps define total and sectoral employment.

The U.S. Census has collected annual County Business Patterns (CBP) data for decades, making them a helpful tool to track the prevalence of information-intensive firms over time in the United States. Isolating the data to the period 2003-2019 allows me to specifically track the rise of information-intensive companies in the United States in the 21st century.

CBP data is particularly useful because it has the most precise level of aggregation at the industry and geographic levels, measuring business activity by not just company headquarters, but by business establishment locations. This fully reflects the inter-regional nature of many businesses today, which do not operate out of a single location but have many establishments located across the country.

In the context of our model, CBP data is helpful in determining both the wage premium $\frac{{w_i}^H}{{w_i}^L}$ between high-skilled and low-skilled workers and the labor ratio $\frac{{L_i}^H}{{L_i}^L}$ between high-skilled and low-skilled workers. More detailed information about how CBP data was cleaned and utilized is available in the Appendix.

4.3 Worker Data

In addition to gathering data at the firm level, we must also gather worker-level data on the amenities that they prefer and the housing costs they incur.

There have been robust debates in the economic literature regarding the types of amenities that workers take into account when moving between different locations. Glaeser et al. (2001) describe four main types of amenities: retail amenities (a variety of goods and services), environmental amenities (aesthetics and weather), crime, and transportation amenities (commuting time/distance). Diamond (2016) expands upon this and defines six types of amenities, including the four from Glaeser et al. (2001), with the addition of schooling and job quality amenities.

However, there is no universally accepted way to measure the value of amenities by MSA. Glaeser et al. (2001) measure amenity indices by MSA by regressing home prices against per capita income and take the residual to be reflective of the quality of local amenities. However, Diamond (2016) found that high- and low-skill workers have different amenity preferences from each other. Given that our research question is predicated on the differential welfare impacts on workers of different skill levels, we must take these findings into account.

I thus take an intermediate approach that combines the simplicity of the Glaeser et al. (2001) approach with Diamond (2016)'s findings of differential amenity preferences. Thus, I calculate an amenity index for high-skilled and low-skilled workers by appropriately regressing the average home value per MSA against per capita income by skill level. The residual of this regression result will thus give us our desired value of ${u_i}^H$ or ${u_i}^L$ accordingly.

In order to calculate this amenity index, however, we must first find appropriate data for housing costs by MSA. I measure housing costs by metropolitan statistical area using 2003-2019 data from the Zillow Home Value Index (ZHVI), which annually measures the typical home value for houses in the 35th to 65th percentile of price for all homes, including single-family residences, condos, and co-ops (Zillow, Inc. (2022)). As opposed to other indices which measure housing costs only through price data of the subset of houses that were sold in a given year, Zillow also incorporates its own market-based estimates of new constructions and existing inventory that is not on the market to calculate the typical home value per MSA, which provides a more comprehensive reflection of the housing market dynamics in a specific MSA. More detailed information about how ZHVI data was cleaned and utilized is available in the Appendix.

Given this paper's focus on welfare outcomes for workers through the channels of the wage and rent gradient, it is natural to wonder why rent data was not used instead of housing data. This is for two primary reasons. First, high-quality data on average rent by MSA has only been available since 2014, and does not include the full scope of MSAs that we wished to study. Second, this limited time range does not provide enough longitudinal data for us to conduct our preferred statistical analyses. However, because rents and home prices move in accordance with one another, and for the purposes of

calculating amenity indices, the distinction between the two measures is not significant, substituting rent data for housing data is acceptable in this instance.

Given these caveats, and based on Glaeser et al. (2001), I generate an equation for our simple regression as follows and accordingly calculate our amenity level indices:

$$ZHVI_{it} = \hat{a} + \hat{b}{w_{it}}^{M} + \hat{\epsilon_{it}}$$

where $ZHVI_{it}$ represents the typical home value in MSA i in year/time t, $\hat{b}$ represents the per-capita income effect by skill level on housing prices in MSA i at year/time t, $H, L \in M$ represents high or low skill level, and residual value $\hat{\epsilon_{it}}$ represents the amenity index in MSA i by skill level at year/time t.

Some stylized results of this regression are displayed in Appendix Tables A.1 and A.2. Notably, there is some divergence between the ranking of different MSAs by amenities available between high- and low-skill workers, especially among the bottom-ranked MSAs, which are very different for each set of workers.

With this derivation of the amenity indices for each MSA as a function of the home value within each MSA, we then have empirical data for the endogenous parameters of our model: w_i and L_i from the firm-level data referenced in Section 4.2, and $\bar{u}_i$ from regressing home value data against wages. Combining these sources of data with each other allows us to generate a novel data set of the MSAs in the United States during this time span with firm and worker-level data collected in aggregate on wages, laborers, amenities, and home values, with relevant distinctions by skill level. The methods of collection of this data, which is further described in the Appendix, thus allow us to answer our research questions of interest.

5 Results

How does information-intensive industry growth affect metropolitan areas and the workers that reside within them? The results of this paper evaluate the empirical data described above while paying heed to the limitations of the aforementioned theoretical model, ultimately shedding light on two questions:

(1) Which metropolitan areas have experienced greater growth of information-intensive firms relative to others?

(2) How do worker outcomes for high- and low-skilled workers differ in these areas with regards to wage, amenities, housing, and welfare, as predicted by our model?

In order to evaluate these effects, I isolate my data set to the 88 MSAs that employed over 25,000 information workers in the year 2003 to prevent our results from being skewed by small sample sizes. Consistent with other econometric panel data analyses, I then utilize a logarithm-transformed fixed effects ordinary least squares model with heteroskedasticity-robust clustered standard errors by MSA. The logarithmic transformation of our predictor and response variables reduces the variance of the residuals in our model and creates more comprehensible coefficients. The fixed effects component of the model takes into account MSA-specific group means and ensures that the correlations between our predictor and response variables are not skewed by MSA-specific factors. Finally, the clustering of standard errors that are heteroskedasticity-robust increases the precision of our estimates.

The results evaluate the effect that increased information employment has on a variety of parameters of interest, including home value, the skilled wage premium, and intrinsic utility by skill level, where f_i is a fixed effect specific to MSA/location i and t represents time since 2003 with dummy variables d to represent year-specific effects:

$$ln(\frac{\hat{w_{it}}^H}{w_{it}{}^L}) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

$$ln(\frac{\hat{u_{it}}^H}{u_{it}{}^L}) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

$$ln(\hat{HomeValue_{it}}) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

This specification allows us to produce unbiased and consistent estimates of the effect of information job growth on each MSA while taking into account both year-specific and MSA-specific effects that recur in the data. Hausman tests of each of the models above confirm the robustness of the fixed effects specification as opposed to a random effects approach. All regression results are robust to heteroskedasticity and the standard errors are clustered by MSA with MSA fixed effects applied. This implies that the ordinary least squares estimates derived are indeed the best linear unbiased estimates of the effect of increased information jobs on wages for low and high-skilled workers within MSAs.

Before delving into the results from these regression specifications, however, I first analyze the data patterns that are visible among the ten highest and lowest information job growth MSAs.

5.1 Information-Intensive Industry Growth

I measure information-intensive industry growth through the annualized percentage change in total information-intensive payroll in each MSA from 2003-2019, which I calculate as follows: $\Delta\$L_i{}^H = \frac{\$L_i^H{}_{2019}}{\$L_i^H{}_{2003}}^{\frac{1}{17}} - 1$.

While there is also data on the rate of change of the number of establishments and of the average wage in each MSA, I use payroll as our measure for growth because it is more indicative of the *scope* of information-intensive firm expansion in a specific MSA. The change in payroll theoretically and empirically better represents the way information-intensive firms are changing metropolitan areas more so than changes to establishment numbers or average wage in the region because it reflects dollars that are being spent on and by laborers within a region. Accordingly, the highest and lowest performing MSAs by information job growth rate from 2003-2019 appear in Table 2.

Notably, many of the highest-growth MSAs are areas that already have high concentrations of information workers – San Jose, San Francisco and Seattle – to name a few. When examining firm growth rates, it's clear that many popular areas for information firms, such as Austin, Raleigh, and Nashville, where firms such as Oracle, Apple, and Amazon respectively have recently announced large offices, appear among the highest-growth areas, whereas many of the low-growth areas are situated in declining industrial centers of the Northeast – Allentown, Bridgeport, New Haven, Springfield, Rochester, and Syracuse are all examples of this.

Furthermore, all of the low-growth MSAs, save for Syracuse, gained total population over this time period, suggesting they lost information jobs despite gaining other types of jobs. This hints at a differential migration pattern between information-heavy MSAs and non-information heavy MSAs. Even MSAs such as San Jose, San Francisco and Seattle, which already had a very high prevalence of information jobs, as can be seen in Appendix Figure A.3, still displayed one of the top growth rates of information jobs despite their high baselines.

Geographically, it is also notable that some of the highest-growth MSAs are located close to the lowest-growth MSAs – Nashville, for example, is located close to Memphis. This implies that geographic location itself and its associated exogenous amenities are not the primary drivers of information employment, or we would observe clusters of low-growth MSAs distinctly from high-growth clusters. Rather, endogenous amenities and characteristics of MSAs are driving the increase in information employment and subsequently the welfare differences hypothesized to occur when information jobs enter a region.

Table 2: Highest and Lowest Information-Intensive Employment Growth Rates as Measured by Annual Payroll Changes, 2003-2019

MSA	High-Growth	%	**MSA**	Low-Growth	%
1.	San Jose, CA	7.7	**1.**	Allentown, PA-NJ	-0.0
2.	Austin, TX	6.8	**2.**	Bridgeport, CT	0.0
3.	San Francisco, CA	6.6	**3.**	New Haven, CT	0.0
4.	Raleigh, NC	6.3	**4.**	Springfield, MA	0.0
5.	Madison, WI	5.8	**5.**	Memphis, TN-MS-AR	0.0
6.	Seattle, WA	5.5	**6.**	Lakeland, FL	0.0
7.	Charleston, SC	5.1	**7.**	Rochester, NY	0.0
8.	Nashville, TN	4.8	**8.**	Jackson, MS	0.0
9.	Grand Rapids, MI	4.7	**9.**	Oxnard, CA	0.0
10.	Huntsville, AL	4.7	**10.**	Syracuse, NY	0.0

Note: This table reports the metropolitan statistical areas (MSAs) with the highest and lowest information-intensive payroll growth rates from 2003 to 2019. These growth rates are measured annualized from changes in payroll from 2003-2019; each MSA listed above on average experienced the percentage change in information-intensive payroll as depicted above every year between 2003 and 2019. Data is of MSAs with greater than 25,000 total information-intensive employees in 2003 to ensure growth rates are not skewed by small sample sizes. Source: *U.S. Census Bureau (2022a)*.

5.2 Wage Effects

I measure the effects of increased information-intensive employment on the skilled wage premium through a logarithmic transformation of our relevant parameters before regressing them against each other. As before, I include only the 88 MSAs with greater than 25,000 information-intensive employees in 2003 to ensure that our data was not skewed by small sample sizes. I also include year dummy variables to adjust for the time-varying effects of wage growth in addition to MSA-specific fixed effects. The year dummies thus captures factors such as inflation, industry-wide payscale changes, and other time-varying effects that may be correlated with changes in information employment. The results of these regressions are displayed for high and low-skilled workers in Table 3, with heteroskedasticity-robust standard errors clustered by MSA.

Overall, I find that *ceteris paribus*, a 1% increase in the information employment workforce of an MSA is associated with a 21.4% increase in the skilled wage premium of that MSA. This significant increase in the skilled wage premium suggests that in areas of high information job growth relative to low-skilled job growth, the discrepancy between high-skilled wages and low-skilled wages grows (the wage premium). This suggests that despite the fact that information job growth is associated with increases in both high- and low-skilled wages as found by Hornbeck and Moretti (2021), wage growth for high-skilled workers is much higher than that of low-skilled workers. Given the fact that high-skilled workers already begin with much higher average wages than low-skilled workers, this increase in the skilled wage premium only exacerbates the disparities between the two sets of workers.

Given the MSAs we identified in Table 2 as areas of high and low information job growth, we can further plot the divergence of wages between these two sets of regions by skill level, as is shown in Figure 4. These results are informative: while high-skill wages in areas of low and high-information growth start off relatively similar to each other in 2003, by 2019, average high-skilled wages in high growth areas are $50,000 greater than their high-skilled counterparts in low growth areas. The year 2010 is a clear divergence point in the trajectories of these two sets of regions, suggesting that much of the effect of the increased skilled wage premium has occurred over the past decade.

Table 3: Effect of Employment on Wage Premium

$$ln(\frac{\hat{w_{it}}^H}{w_{it}{}^L}) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

	Dependent variable:
	$ln(\frac{w_i{}^H}{w_i{}^L})$
$ln(\frac{L_i{}^H}{L_i{}^L})$	0.214***
	(0.035)
Observations	1,443
Within R^2	0.388
Adjusted R^2	0.883
MSA Fixed Effects	Yes
Year Dummies	Yes
Note:	*p<0.1; **p<0.05; ***p<0.01

This table reports estimates of the effect of a rise in the high-skilled labor ratio on the skilled wage premium. The column reports coefficients from a regression of log skilled wage premium on log labor ratio with MSA fixed effects and year dummies. The results are from a model that includes clustered standard errors by MSA, with standard errors reported in parentheses.

Figure 4: Wage Differences in Areas of Low and High Information Sector Growth

Note: This figure reports differences in average wage by skill level in areas of high and low information job growth. High-growth and low-growth MSAs are defined in Table 2. Source: *U.S. Census Bureau (2022a)*.

This implies that for high-skilled workers, living in a high information growth region has not only increased the level of their earnings but also the trajectory of their earnings growth. On the other hand, while low-skill wages in areas of high information job growth are at a higher level than in low-growth regions over time, the trajectory of the two regions are identical. This, along with the regression results, suggests that the addition of information jobs into MSAs increases the skilled wage premium, which comes at a detriment to low-skilled workers who, over time, have become poorer compared to their high-skilled counterparts. Overall, these findings provide evidence against the idea that there exists wage convergence across skill levels, instead implying that information-heavy regions see greater productivity (and hence, wages) across skill levels.

These results suggest that the rising tide of information jobs can exacerbate the differences in workers' take-home pay, even if it is associated with increased wages for both sets of workers. We see empirically from both our regression results and the figure that an increase in high-skilled labor relative to low-skilled labor $\frac{{L_i}^H}{{L_i}^L}$ is associated with a significant increase in the skilled wage premium $\frac{{w_i}^H}{{w_i}^L}$.

5.3 Amenity Effects

As referenced in Section 4.3, I calculated each MSA's amenity index by taking the residuals of the regression of average wage versus home value. These residuals then reflect the intrinsic value of living in a location that is not reflected in the level of wages paid to workers in that location. However, because we wish to identify the differences in amenity value of locations between workers of different skill levels, I run two separate regressions, which generate amenity indices for each location for the two sets of workers we are considering: high-skilled and low-skilled workers. This reflects Diamond (2016)'s empirical observations of differential amenity preferences among many high-skilled workers (who tend to be college-educated) as opposed to low-skilled workers (who tend not to be college-educated).

Given these amenity indices for low- and high-skilled workers, the next step is to ascertain whether the increased presence of high-skilled employees relative to low-skilled employees would affect the amenity indices for either high- or low-skilled workers. In

other words, would increased information employment in an MSA affect the intrinsic utility value of that location for either low- or high-skilled workers? As before, I include both MSA fixed effects and year dummies in order to adjust for the time-varying effects of changing amenity values. These year dummy variables may capture factors such as year-over-year trends in the way workers value MSAs that may be correlated with changes in information employment. The results of these regressions are displayed in Table 4.

Overall, we find that increases in the information workforce relative to the low-skilled workforce significantly affect the intrinsic utility value of an MSA: a 1% increase in the skilled labor ratio of an MSA is associated with a 7.5% increase in the intrinsic utility value of that MSA for high-skilled workers relative to low-skilled workers. This suggests that increased information employment is associated with more attractive locations for high-skilled workers and suggests a degree of residential sorting that is in line with findings by Diamond (2016): high-skilled workers have higher utility values for MSAs with a greater prevalence of high-skilled workers. This residential sorting pattern, if taken to its extreme, then suggests the formation of cities that essentially function as enclaves for only high-skilled workers, who are both willing and able to pay for the amenities of the location given their high wages. This runs counter to the idea under welfare equalization that locations must compensate for high wages by having a lower level of amenities; rather, for high-skilled workers, they are able to benefit from high wages and a high level of amenities, yet again suggesting that welfare equalization does not hold, even within workers of the same skill level.

While there is a significant association between utility values of MSAs and information employment, information employment in isolation is not fully explanatory of these high utility values. Much of this is likely due to the exogenous factors that influence utility that are not included in our regression specification, such as climate, precipitation, days of sunlight, etc. Looking at Appendix Tables A.1 and A.2 show that the top 10 MSAs by amenity levels tend to be sunny, coastal cities, whereas the bottom 10 MSAs by amenity level tend to be inland and colder. Amenity values thus are not explained by the presence of information sector workers, but seem to be associated with them.

Our findings suggest that increased information employment is associated with an increased skilled wage premium and an increased desirability of that location for high-

Table 4: Effect of Employment on Skilled Amenity Ratio

$$ln(\frac{\hat{u_{it}}^H}{u_{it}{}^L}) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

	Dependent variable:
	$ln(\frac{u_i{}^H}{u_i{}^L})$
$ln(\frac{L_i{}^H}{L_i{}^L})$	0.075***
	(0.029)
Observations	1,443
Within R^2	0.012
Adjusted R^2	0.445
MSA Fixed Effects	Yes
Year Dummies	Yes
Note:	*p<0.1; **p<0.05; ***p<0.01

This table reports estimates of the effect of a rise in the high-skilled labor ratio on the amenity value of an MSA for high-skilled workers relative to low-skilled workers. The column reports coefficients from a regression of log amenity ratio on log labor ratio with MSA fixed effects and year dummies. The results are from a model that includes clustered standard errors by MSA, with standard errors reported in parentheses.

skilled workers relative to low-skilled workers. In short, it seems that information job growth is associated with a divergence in both the wages and the intrinsic utility values of MSAs between low- and high-skilled workers. High-skilled workers seem to value being in close proximity to other high-skilled workers, as was found by Diamond (2016), and also earn more money when they are.

I next consider the effect of information job growth on home values within an MSA. Increasingly, housing makes up the largest percentage of consumers' expenditure and the affordability of housing is a primary driver of migration into and out of specific regions. Ascertaining the effect of increased information employment on home values provides additional insight into the effect of information job growth on welfare outcomes for high- and low-skilled workers.

5.4 Housing Effects

Concerns about gentrification and rising home values pricing out low-skilled workers have been raised numerous times in the context of information firms locating in cities. To empirically test this assertion, as before, I include dummy variables for the year in order to adjust for the time-varying effects of changing home values to reflect inflation and other year-specific housing market trends. I also include MSA-level fixed effects to account for the variations in price between different MSAs in order to ensure our estimates are unbiased. The results of these regressions are displayed in Table 5.

Overall, we find that *ceteris paribus*, a 1% increase in the information employment ratio of an MSA is associated with an increase in the average home value of a location of approximately 23.4%. Increased information employment relative to low-skilled employment is thus associated with higher home values, which affects both low- and high-skilled workers in a given MSA. Given that low-skilled workers are less likely to be homeowners and will not benefit from the rising equity value of their homes (and correlated rising rents), for low-skilled workers in areas of high information job growth, this home value increase harms them disproportionately.

Table 5: Effect of Employment on Home Values

$$ln(HomeValue) = \hat{\beta}_1 * ln(\frac{L_{it}{}^H}{L_{it}{}^L}) + d_t + f_i + \epsilon_{it}$$

	Dependent variable:
	$ln(HomeValue)$
$ln(\frac{L_i{}^H}{L_i{}^L})$	0.234***
	(0.092)
Observations	1,443
Within R^2	0.643
Adjusted R^2	0.962
MSA Fixed Effects	Yes
Year Dummies	Yes
Note:	*p<0.1; **p<0.05; ***p<0.01

This table reports estimates of the effect of a rise in the high-skilled labor ratio on home values in an MSA. The column reports coefficients from a regression of log home value on log labor ratio with MSA fixed effects and year dummies. The results are from a model that includes clustered standard errors by MSA, with standard errors reported in parentheses.

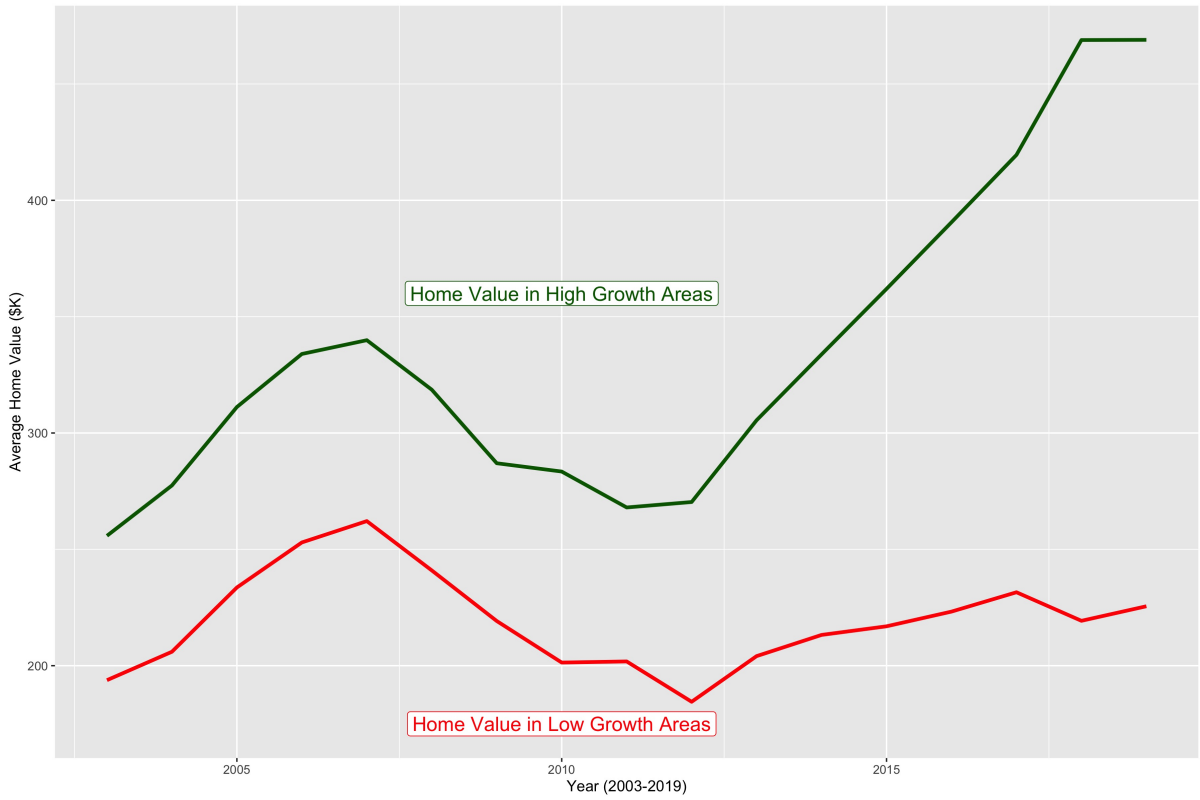
While information employment does not entirely explain the rising home values of MSAs with high information job growth, there is a strong and significant association between information employment and home value, again suggesting that increased information employment is correlated with deepening inequality between high and low-skilled workers.

Home values in areas of high information job growth have significantly diverged from their low growth counterparts. Although we observe from Figure 5 that areas with high information job growth start off with higher average home values by approximately $50,000, these areas also witness larger increases in home value over time. In 2019, high information job growth areas had average home values over $225,000 greater than their low growth counterparts. On the other hand, for areas with low information job growth, home value has actually declined over time and has not recovered to its peak before the housing market crash of 2008-2009. A divergence point is evident around the year 2012, when home values in high growth areas begin to increase at a much more rapid pace than their low growth counterparts. Overall, this pattern indicates house prices are growing in a very spatially unequal manner, creating an affordability crisis in areas of high information job growth for not just low-skilled workers, but workers in those industries that are neither low nor high-skilled (as seen in Figure 2).

These results emphasize the importance of considering the effects of the limited urban housing stock in consumer welfare. Areas of high information job growth have seen significant increases in housing costs, which in turn pressures low-skilled residents with lower wages to out-migrate from these regions. This is in line with Diamond (2016)'s findings that workers of different skill levels have increasingly sorted into different cities, resulting in a number of metropolitan areas with abnormally low or high ratios of high- to low-skilled workers, as can be seen in Appendix Figure A.3.

For policymakers, this also suggests the importance of ensuring cities remain accessible to workers of all skill levels, primarily by replenishing the existing housing stock in urban areas and increasing the development of affordable housing for low-skilled workers.

Figure 5: Home Value Differences in Areas of Low/High Information Growth



Note: This figure reports differences in average home value in areas of high and low information job growth. High-growth and low-growth MSAs are defined in Table 2. Source: *Zillow, Inc. (2022)*.

5.5 Welfare Effects and Evaluation of the Model

Given our results, we found that the growth of information employment was associated with significant increases to the skilled wage premium ($\frac{w_i{}^H}{w_i{}^L}$), intrinsic amenity ratio of locations ($\frac{\bar{u_i}^H}{u_i{}^L}$), and home values.

Based on the model derived from Giannone (2011) and Allen and Arkolakis (2014), welfare discrepancies $\frac{W_i{}^H}{W_i{}^L}$ are calculated as $\frac{w_i{}^H}{w_i{}^L}\left(\frac{L_i{}^H}{L_i{}^L}\right)^{-\beta}\left(\frac{\bar{u}_i^H}{\bar{u}_i^L}\right)$. Given that a rise in $L_i{}^H$ relative to $L_i{}^L$ is associated with significant increases in both the wage premium $\frac{w_i{}^H}{w_i{}^L}$ and the amenity ratio $\frac{\bar{u_i}^H}{u_i{}^L}$, the model suggests that (given small enough β dispersion forces), information job proliferation actually *increases* the welfare discrepancies between high-skilled and low-skilled workers, implying that information job proliferation is a force that may further contribute to urban inequality within cities. These welfare discrepancies can also be restated as $\frac{W_i{}^H}{W_i{}^L} = \left(\frac{\bar{A}_i^H}{\bar{A}_i^L}\right)^{\frac{\rho-1}{\rho}}\left(\frac{L_i{}^H}{L_i{}^L}\right)^{\frac{-(1+\alpha)}{\rho}-\beta}\left(\frac{\bar{u}_i^H}{\bar{u}_i^L}\right)$. The results thus also imply that the value of ρ, a parameter that reflects the elasticity of labor substitution between high-skilled and low-skilled workers (which thus measures the degree of skill-biased technological change), is greater than 1, and that the strength of dispersion forces β do not outweigh the positive effects of wage and productivity growth for high-skilled workers' welfare. In short, because high-skilled workers become more essential to firms' production processes (through an increase in ρ over time), low-skilled workers do not see nearly as many gains in welfare. Since this model incorporates skill heterogeneity, its mechanism implies that welfare discrepancies are specifically exacerbated through the channel of the skilled wage premium and amenity ratio, and the channel through which spatial equilibrium is achieved is through the in-migration of high-skilled workers and out-migration of low-skilled workers, respectively.

However, the model as it stands does not directly incorporate home values into its calculation of welfare, which is a shortcoming. In modelling a perfect world with unlimited housing stock, information jobs may still be welfare-enhancing for low-skilled workers due to the increases in low-skilled wages associated with information job growth. But because of the limited housing stock available in cities and the significant association between rising home value and information jobs, low-wage workers in these areas of high information job growth are increasingly limited in their options of where to live, as was

found by Hsieh and Moretti (2019).

The results as compared to the model also provide intriguing perspective for policymakers and economists. The lack of significant wage growth for low-skilled workers stemming from information job growth runs counter to the idea that information-intensive firm entrance increases wage for all workers through the channel of productivity spillovers, as was observed by Hornbeck and Moretti (2021) to be the case in the 1980s. Instead, it implies that there has been a decoupling of productivity and wages for certain sets of workers, whereby improvements in labor productivity are no longer being internalized into the wage gradient. This aligns with previous findings from Fleck et al. (2011) that have shown a decreasing share of national income going towards labor and instead more going towards capital, despite increased labor productivity. Traditional economic assumptions that have coupled wage and productivity together should be further scrutinized and re-evaluated, especially for low-skilled workers. This is especially salient in the context of the COVID-19 pandemic, where remote work options for high-skilled workers has further enhanced their productivity, but has not been option for low-skilled workers.

Overall, the empirical data as applied to the theoretical model incorporating skill-biased technological change seems to suggest that information job growth exacerbates the welfare discrepancies between workers of high and low skill levels and that this effect has increased over time. However, since the model does not fully incorporate the limited housing stock, which further restricts the population of an MSA, it may even underestimate the true extent of these welfare differences across time.

6 Discussion and Conclusion

Information-intensive industry growth has reshaped a number of metropolitan areas in recent years. In order to understand how these changes affected workers of high- and low-skill levels, I employed a fixed effects model on panel data for metropolitan areas from 2003-2019 and compared the results to a canonical urban economic model that is adjusted for skill-biased technological change. These methods of analysis allowed for compelling answers as to how information-intensive firms affect worker outcomes in areas

of high information job growth and which metropolitan areas have been most affected by these trends.

Information-intensive job growth was associated with significant increases in the skilled wage premium, amenity ratio, and home values of MSAs, taking into account MSA-level fixed effects. Overall, I estimate that a 1% increase in information employment in an MSA relative to low-skill employment is associated with a 21.4% increase in the skilled wage premium, a 7.5% increase in the intrinsic utility ratio of an MSA, and a 23.4% increase in home value. Due to the comparative lack of wage growth for low-skilled workers over time as opposed to that of high-skilled workers, these home value increases, given the limited housing stock of metropolitan areas, disproportionately burden low-skilled workers.

These findings raise important questions for policymakers with regards to information firms' power to influence governments through procuring attractive economic development subsidies, tax breaks, and other monetary incentives that may end up harming low-skilled workers and exacerbating the housing affordability crises in cities today. Given the fact that many cities and states are actively *subsidizing* information companies such as Amazon to locate in their regions, the ethics of using taxpayer dollars on projects associated with increasing housing costs and increased income inequality through the higher skilled wage premium is questionable. Re-evaluating the need to subsidize such projects, or including affordable housing requirements attached with such large projects, should be considered to reduce the impact of these projects on existing residents.

There are, however, limitations to the research conducted in this paper. The theoretical model incorporating skill-biased technological change does not fully incorporate housing into its measure of consumer welfare and, given our results, may actually underestimate the welfare discrepancies between high-skilled and low-skilled workers. For example, Hsieh and Moretti (2019) find that cities with high information-driven productivity have also been more prone to adopting local housing constraints that have limited the supply of housing. In the future, canonical urban economic models should thus be amended to reflect the impact of housing as well.

While the aggregation of firm and worker-level data at the metropolitan statisti-

cal area (MSA) level is helpful in analyzing broad trends, in the future, neighborhood-level data would be even more beneficial in allowing for the use of quasi-experimental methods to determine causal links between information firm entrance and its effects on wages, amenities, and rents.

Overall, not only does spatial inequality *within* metropolitan areas seem to increase with greater information-intensive firm presence, but it also does *between* metropolitan areas. This distinction between metropolitan areas with high information job presence versus those without significant information job presence only exacerbates Moretti (2012)'s findings regarding the "Great Divergence" between high-skilled and low-skilled workers. Already existing class-based segregation within and between metropolitan areas will increase without further intervention.

Though wage inequality will continue to persist because of the skilled wage premium, the sudden and sharp rise of housing costs in cities with high information job presence is indicative of the need for both policymakers and economists to update their models and policies to reflect the patterns of the city as it enters the information age. As Figure 1 demonstrates, there is no going back on broader macroeconomic patterns of information-intensive industry growth. The challenge now is to ascertain and implement distinct ways to encourage information-intensive firm presence without contributing to urban inequality.

As the information age continues to accelerate and information firms expand their presence in cities, the composition of our cities will inevitably change as well. The results suggest that the rising tide of information jobs is associated with meaningful increases in the skilled wage premium, amenity ratio, and housing costs, ultimately increasing spatial inequality through differential welfare impacts on different sets of skilled workers. Taken to its most extreme form, these patterns suggest that in the cities with the highest levels of productivity, only wealthier individuals will be willing and able to pay to live there. This suggests the need for state intervention in the form of place-based policies to improve existing housing stock and build new affordable housing in areas of high information job growth rather than limiting the supply of housing. Without such interventions, we may see metropolitan areas with high information employment presence increasingly pushing out low-skilled workers.

References

Allen, T. and C. Arkolakis (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics 129*(3), 1085–1140. https://doi.org/10.1093/qje/qju016.

Amazon.com, Inc. (2017). Amazon HQ2 RFP. https://images-na.ssl-images-amazon.com/images/G/01/Anything/test/images/usa/RFP_3._V516043504_.pdf.

Amazon.com, Inc. (2018). Amazon Selects New York City and Northern Virginia for New Headquarters. https://press.aboutamazon.com/news-releases/news-release-details/amazon-selects-new-york-city-and-northern-virginia-new.

Amazon.com, Inc. (2019). Update on plans for New York City headquarters. https://www.aboutamazon.com/news/company-news/update-on-plans-for-new-york-city-headquarters.

Boustan, L., D. Bunten, and O. Hearey (2018). Urbanization in the United States, 1800-2000. *The Oxford Handbook of American Economic History*. https://www.nber.org/system/files/working_papers/w19041/w19041.pdf.

Cronon, W. (1992). *Nature's Metropolis*. New York: W.W. Norton.

Davis, D. R. and J. I. Dingel (2019). A Spatial Knowledge Economy. *American Economic Review 109*(1), 153–170. https://doi.org/10.1257/aer.20130249.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review 106*(3), 479–524. http://dx.doi.org/10.1257/aer.20131706.

Eckert, F., S. Ganapati, and C. Walsh (2020). Skilled Scalable Services: The New Urban Bias in Economic Growth. *Federal Reserve Bank of Minneapolis* (25). https://doi.org/10.21034/iwp.25.

Fleck, S., J. Glaser, and S. Sprague (2011). The compensation-productivity gap: a visual

essay. *U.S. Bureau of Labor Statistics Monthly Labor Review*. https://www.bls.gov/opub/mlr/2011/01/art3full.pdf.

Giannone, E. (2011). Skill-biased Technical Change and Regional Convergence. https://crei.cat/wp-content/uploads/2022/01/JMP_March2021.pdf.

Gibson, C. (1998). Population of the 100 Largest Cities and Other Urban Places In The United States: 1790 to 1990. Working Paper POP-WP027, U.S. Census Bureau. https://www.census.gov/library/working-papers/1998/demo/POP-twps0027.html.

Glaeser, E. L., J. Kolko, and A. Saiz (2001). Consumer City. *Journal of Economic Geography 1*(1), 27–50. https://www.nber.org/system/files/working_papers/w7790/w7790.pdf.

Hornbeck, R. and E. Moretti (2021). Estimating Who Benefits From Productivity Growth: Local and Distant Effects of City Productivity Growth on Wages, Rents, and Inequality. https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/b/2217/files/2021/10/TFP_HM_Oct2021.pdf.

Hsieh, C.-T. and E. Moretti (2019). Housing Constraints and Spatial Misallocation. *American Economic Journal 11*(2), 1–39. https://doi.org/10.1257/mac.20170388.

Moretti, E. (2012). *The New Geography of Jobs*. New York: First Mariner Books.

O'Connell, J. (2017). Amazon receives 238 proposals for second headquarters. *The Washington Post*. https://www.washingtonpost.com/news/business/wp/2017/10/23/amazon-receives-238-proposals-for-second-headquarters/.

Roback, J. (1982). Wages, Rents, and the Quality of Life. *Journal of Political Economy 90*(6), 1257–1278. https://doi.org/10.1016/j.jue.2004.04.001.

Rosen, S. (1979). Wage-based Indexes of Urban Quality of Life. *Current Issues in Urban Economics*, 74–104.

Sinai, T. and J. Waldfogel (2004). Geography and the Internet: is the Internet a substitute or a complement for cities? *Journal of Urban Economics 56*, 1–24. https://doi.org/10.1016/j.jue.2004.04.001.

U.S. Bureau of Economic Analysis (2022). Gdp by Industry. Dataset. https://www.bea.gov/data/gdp/gdp-industry.

U.S. Census Bureau (2010). Cartographic Boundary Files - Shapefile. Dataset. https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2010.html.

U.S. Census Bureau (2022a). County Business Patterns (CBP) Datasets. Dataset. https://www.census.gov/programs-surveys/cbp/data/datasets.html.

U.S. Census Bureau (2022b). North American Industry Classification System. https://www.census.gov/naics/?58967?yearbck=2022.

Zillow, Inc. (2022). Zillow Housing Data. Dataset. https://www.zillow.com/research/data/.

A Appendix

This Appendix includes further details on how I collected, cleaned, and processed my data, in addition to a number of other explanatory figures and tables that shed light on the paper's results and conclusions.

A.1 Data Cleaning

Geospatial Data

The U.S. Census Bureau makes geographic shapefiles of the U.S. states and metropolitan statistical areas available on its website at https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.2010.html. I downloaded these shapefiles and loaded them into the geographic information system (GIS) software of my choice, which is QGIS 3.20. By merging the delimited text layers of data described below onto these shapefiles, I am able to produce maps and figures that display variables in a geospatial format.

Firm Data

The U.S. Census Bureau makes its annual County Business Patterns (CBP) data available as raw text files on its website at https://www.census.gov/programs-surveys/cbp/data/datasets.html. Data aggregated by metropolitan statistical areas is available from 1993 onwards, but the data from 1993 to 2003 uses obsolete MSA categorizations and codes that cannot easily be compared to post-2003 MSA-level data. Hence, I isolate my analysis to data from 2003 onwards.

I converted each raw text file to comma-separated value files and selected for our NAICS Codes of Interest, as defined in Section 1.2. Accordingly, using this data, I was able to calculate the average wage, wage premium, and labor ratio per MSA.

Worker Data

Zillow publishes monthly data for the Zillow Home Value Index (ZHVI) on its website at https://www.zillow.com/research/data/. Data aggregated by metropolitan statistical area is available since 1996. However, given the fact that our firm-specific data is measured annually and is only usable since 2003, we isolate our analysis to ZHVI observations from January 2003 until December 2019. We also average the 12 months of each calendar year to create a single, annual measure of typical home value per MSA per year. This ensures all of the data we use is of the same temporal frequency.

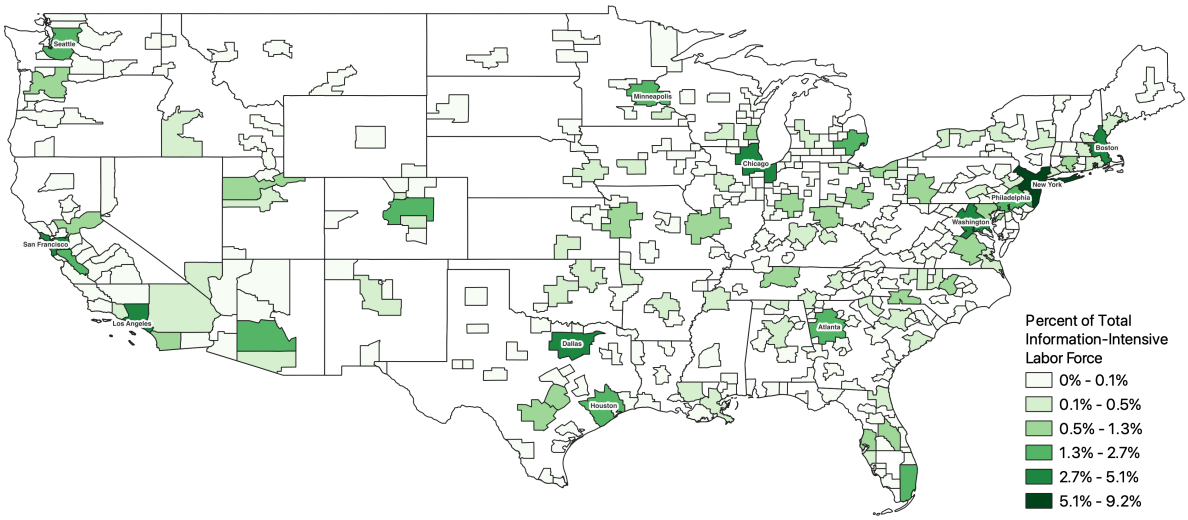
ZHVI data is available in comma-separated value format, but Zillow does not identify MSAs through the same FIPS codes as the U.S. Census Bureau does, instead generating its own region identification numbers. I thus utilize a crosswalk file between Zillow identifiers and the Census's MSA identifiers, found at http://files.zillowstatic.com/research/public/CountyCrossWalk_Zillow.csv, in order to make sure our data can match up with Census region identifiers and accordingly be mapped onto our geographic shapefiles for easier visual analysis.

Based on the residuals derived from the regression model described in Section 4.3, I calculate a skill-based amenity index for each MSA by year. This regression utilizes both the ZHVI and CBP data described above.

A.2 Figures

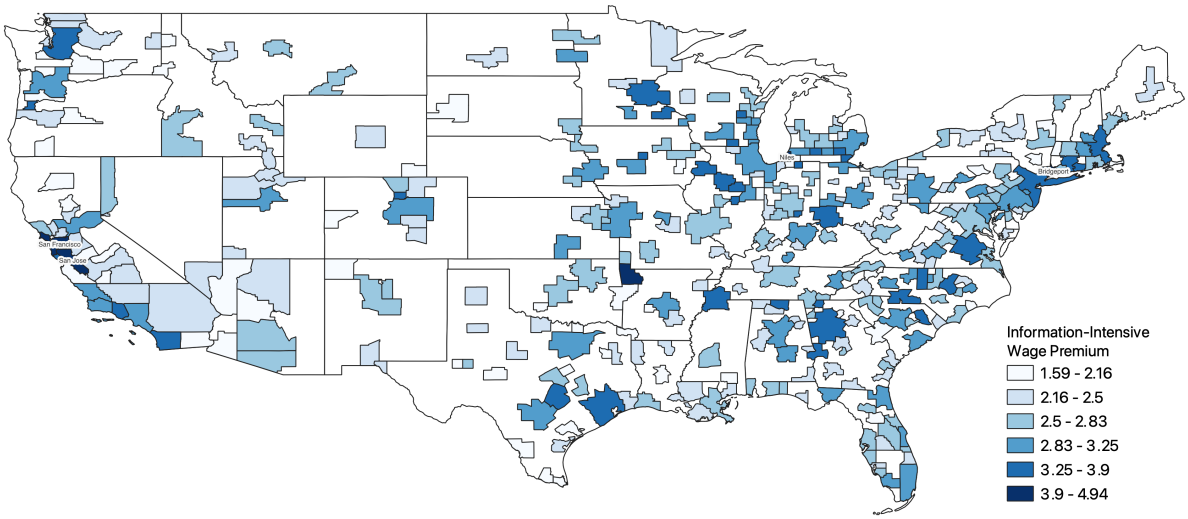
The following figures provide a geospatial visualization of the agglomeration of information-intensive firms in specific MSAs (Figure A.1), the skilled wage premium (Figure A.2), the ratio of high-skilled to low-skilled workers (Figure A.3), and average home value (Figure A.4).

Figure A.1: Percent of Total Information-Intensive Labor Force by Metropolitan Statistical Area, 2018



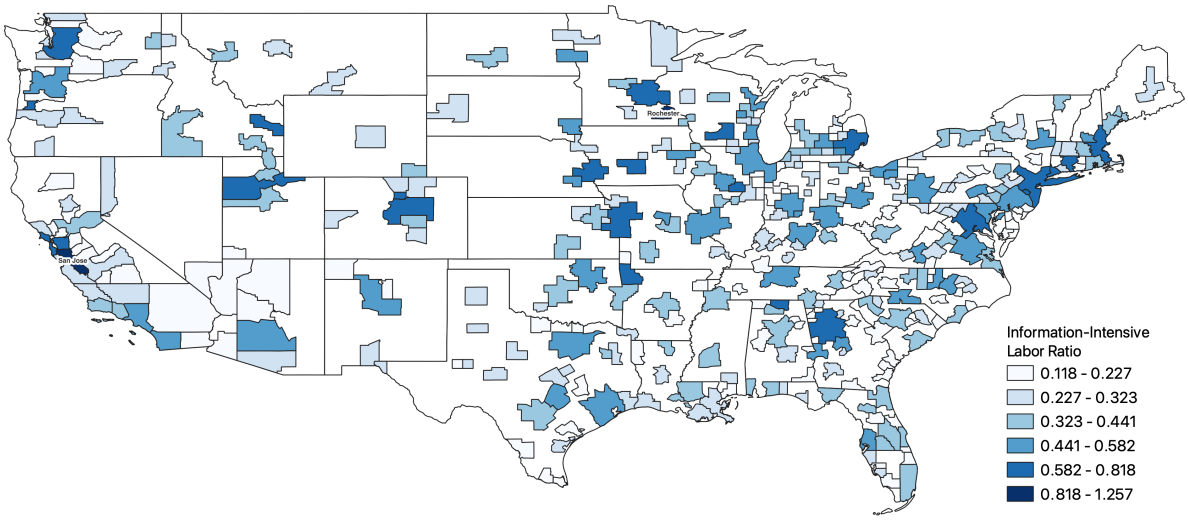
Note: This figure reports the percentage of the total information-intensive workforce located in each metropolitan statistical area across the country. Each colored area represents a distinct metropolitan statistical area (MSA), and white areas are not located within any MSA. MSAs with over 2% of nationwide information-intensive laborers are labeled. Geographic shapefiles of each MSA overlaid upon the U.S. states are both derived from U.S. Census Bureau (2010). Legend colors and distinctions are determined through a Jenks natural breaks optimization. Source: *U.S. Census Bureau (2022a)*.

Figure A.2: Information-Intensive Wage Premium by Metropolitan Statistical Area, 2019



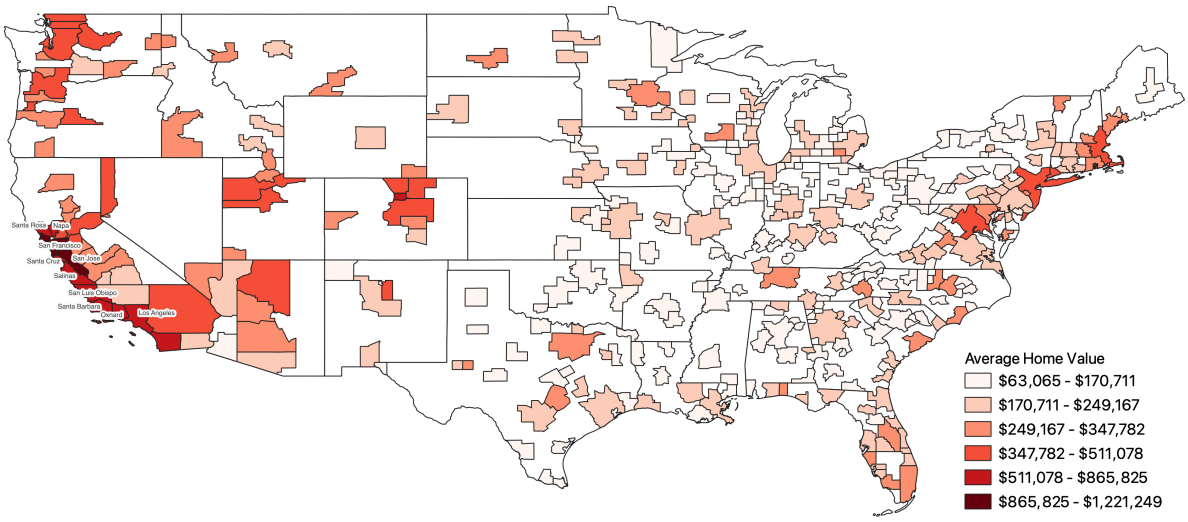
Note: This figure reports the skilled wage premium in each metropolitan statistical area across the country. Each colored area represents a distinct metropolitan statistical area (MSA), and white areas are not located within any MSA. MSAs with a skilled wage premium of over 4.2, in which the average information/high-skilled worker makes over 4.2 times the average low-skilled worker, are labeled. Geographic shapefiles of each MSA overlaid upon the U.S. states are both derived from U.S. Census Bureau (2010). Legend colors and distinctions are determined through a Jenks natural breaks optimization. Source: *U.S. Census Bureau (2022a)*.

Figure A.3: Information-Intensive Labor Ratio by Metropolitan Statistical Area, 2019



Note: This figure reports the ratio of high-skilled to low-skilled laborers in each metropolitan statistical area across the country. Each colored area represents a distinct metropolitan statistical area (MSA), and white areas are not located within any MSA. Metropolitan statistical areas (MSAs) with a labor ratio of at least 1, in which the number of high-skilled, information-intensive workers exceed the number of low-skilled workers, are labeled. Geographic shapefiles of each MSA overlaid upon the U.S. states are both derived from U.S. Census Bureau (2010). Legend colors and distinctions are determined through a Jenks natural breaks optimization. Source: *U.S. Census Bureau (2022a)*.

Figure A.4: Average House Price by Metropolitan Statistical Area, 2019



Note: This figure reports the average home value in each metropolitan statistical area across the country. Each colored area represents a distinct metropolitan statistical area (MSA), and white areas are not located within any MSA. Metropolitan statistical areas (MSAs) with average home prices of at least $600,000 are labeled. Data is from the Zillow Home Value Index, which measures the home value for houses in the 35th to 65th percentile of price for all homes, including single-family residences, condos, and co-ops. Geographic shapefiles of each MSA overlaid upon the U.S. states are both derived from U.S. Census Bureau (2010). Legend colors and distinctions are determined through a Jenks natural breaks optimization. Source: *Zillow, Inc. (2022)*.

A.3 Tables

The following tables are generated from the methods described in Section 4.3, whereby amenity indices are computed from combining the insights of Glaeser et al. (2001) and Diamond (2016). These results are compiled from the residuals of the following regression:

$$ZHVI_{it} = \hat{a} + \hat{b}{w_{it}}^{M} + \hat{\epsilon_{it}}$$

where $ZHVI_{it}$ represents the typical home value in MSA i in year/time t, $\hat{b}$ represents the per-capita income effect by skill level on housing prices in MSA i at year/time t, $H, L \in M$ represents high or low skill level, and residual value $\hat{\epsilon_{it}}$ represents the amenity index in MSA i by skill level at year/time t.

Table A.1: Top 10 MSAs per Skill Level by Amenity Index, 2019

MSA	High-Skill	Low-Skill
1.	San Francisco	San Jose
2.	San Jose	San Francisco
3.	Santa Cruz	Oxnard
4.	Los Angeles	Santa Cruz
5.	San Diego	Boulder, CO
6.	Oxnard	San Diego
7.	Riverside	Los Angeles
8.	Boulder, CO	Sacramento
9.	Salt Lake City, UT	Riverside
10.	Sacramento	Raleigh, NC

Note: This table reports the most desirable metropolitan statistical areas for workers of each skill level as measured by the residuals of a regression of home value on average wage. Unless otherwise indicated, all MSAs above are located in the state of California. Amenity indices are calculated by taking the residuals from regressing housing prices against per-capita income per skill level per MSA. Per-capita income is measured in nominal terms. Data on house prices is derived from an annual average of the Zillow Home Value Index (Zillow, Inc. (2022)).

Table A.2: Bottom 10 MSAs per Skill Level by Amenity Index, 2019

MSA	High-Skill	Low-Skill
1.	Bridgeport, CT	Washington, DC-MD-VA-WV
2.	Houston, TX	Cleveland, OH
3.	Hartford, CT	Las Vegas, NV
4.	Memphis, TN-MS-AR	Chicago, IL-IN-WI
5.	Chicago, IL-IN-WI	Detroit, MI
6.	Trenton, NJ	Indianapolis, IN
7.	Pittsburgh, PA	New Orleans, LA
8.	Detroit, MI	Albany, NY
9.	Dayton, OH	Pittsburgh, PA
10.	Cincinnati, OH-KY-IN	Memphis, TN-MS-AR

Note: This table reports the least desirable metropolitan statistical areas for workers of each skill level as measured by the residuals of a regression of home value on average wage. Amenity indices are calculated by taking the residuals from regressing housing prices against per-capita income per skill level per MSA. Per-capita income is measured in nominal terms. Data on house prices is derived from an annual average of the Zillow Home Value Index (Zillow, Inc. (2022)).