HARVARD | BUSINESS | SCHOOL

Digital Labor Market Inequality and the Decline of IT Exceptionalism

Ruiqing Cao
Shane Greenstein

Working Paper 21-019

Digital Labor Market Inequality and the Decline of IT Exceptionalism

Ruiqing Cao
National Bureau of Economic Research

Shane Greenstein
Harvard Business School

Working Paper 21-019

Funding for this research was provided in part by Harvard Business School and Sloan Foundation through the National Bureau of Economic Research.

Digital Labor Market Inequality and the Decline of IT Exceptionalism*

Ruiqing Cao † Shane Greenstein ‡

July 2020

Abstract

Several decades of expansion in digital communications, web commerce, and online distribution have altered the U.S. labor market for IT workers. We characterize the shifts in regional IT labor markets from 2000 to 2018, and find that IT wage growth did not follow an exceptional pattern compared to broader STEM labor market trends. Digital wage inequality increased, almost entirely due to rising local premiums in a few urban metropolises, where wage spreads became narrower than elsewhere. The supply of college-educated workers accounted for a substantial share of the total wage difference between IT labor markets in top locations and other cities. Agglomeration and IT innovation explained a notably larger fraction of the top-location wage premium in more recent years.

*We thank the Sloan Foundation for financial support through the National Bureau of Economic Research. We thank Chris Forman for providing processed patents data. We are grateful to Monica Farid, Chris Forman, Emma Harrington, Maggie Kelleher, Chris Stanton, and Alice Wu for helpful comments. All errors are our own.

†National Bureau of Economic Research. Email: rcao@nber.org

‡Harvard Business School. Email: sgreenstein@hbs.edu

1 Introduction

Beginning with the growth of the personal computer in the 1980s, economic analysis has considered why IT wages command a premium. Theories of skill-biased technical change offer an explanation: the premium to wages among IT workers results from the demand for the additional and rare skills required to employ frontier IT for purposes with high market value (Autor, Katz and Krueger, 1998; Autor, Levy and Murnane, 2003; Goldin and Katz, 2007; Autor, Katz and Kearney, 2006; Autor et al., 2019; Goldin, Katz et al., 2020). Viewed through this lens, the uneven deployment of the Internet in the 1990s placed additional pressure on IT wages in some locations, and offers one explanation for why the distribution of regional wages became less equal (Forman, Goldfarb and Greenstein, 2002, 2005).

After the turn of the millennium, the pressures leading to skill-biased technical change persisted. U.S. businesses continued to make substantial investments in networking infrastructure and enterprise IT and this class of assets grew faster than any other in the nationwide economy (Byrne and Corrado, 2019). Technical advances, such as broadband Internet access, Web 2.0 and 3.0, web-software to support video advertising, streaming, smartphone applications, data center hardware and software to enable big data analytics, machine learning, and cloud technology, diffused across many industries and locations (Greenstein, 2020). These advances created visible changes in the skill requirements for IT labor. Training changed for software engineers and architects, system administrators and developers, information analysts, and support specialists, adding new software language staples to frontier skill requirements, such as JavaScript, Nginx, Hadoop, and neural networks (Tambe, 2014; Horton and Tambe, 2019).

Did the experience after the millennium heighten or lessen regional inequality, or exacerbate the response to skill-biased technical change? A general answer is not obvious from descriptive data and existing studies. Overall, the new frontier created different degrees of changes in the local returns to skill, but IT is only one of many occupations that require skills. Some urban environments broadly supported higher returns to high-skilled labor (Moretti, 2013; Baum-Snow and Pavan, 2013; Baum-Snow, Freedman and Pavan, 2018), and to frontier IT production and usage in a small number of urban clusters (Tambe and Hitt, 2012; Forman, Goldfarb and Greenstein, 2012). Whether that led to a higher or lower wage premium is an empirical question. So too is whether the results differed substantially across different types of skilled labor other than IT.

This study compares variance in regional IT wage premiums with other skilled labor markets and analyzes the causes of divergence and convergence between them. If skill-biased technical change shapes IT labor wages, does it also shape the wage premium in Science, Technology, Engineering, and Mathematics (STEM) and do these operate to a comparatively similar degree across regions? If the premium in information technology is exceptional, what factors explain the divergence from trends in STEM labor markets in some locations and not others? If unexceptional, what factors shapes all skilled wages?

To motivate these comparisons, consider the regional variance in a few illustrative jobs. As shown by the Occupational Employment Statistics, IT occupations account for 40-50% of the overall STEM employment of 9.7

million workers in 2018 (we ask the reader to momentarily defer questions about definitions and details, which we address in the paper). The average wages of computer and information research scientists in the locations with the highest wages in 2018 were $167,990 in Santa Clara (Silicon) Valley and $140,660 in San Francisco. That compares to $117,260 in the Indianapolis. San Francisco and Santa Clara (Silicon) Valley account for 6.7% of IT employment in the United States. At the same time, the average wages of biochemists were $115,070 in San Francisco and $105,850 in Santa Clara (Silicon) Valley, compared to $118,790 in Indianapolis. In that case, San Francisco and Santa Clara (Silicon) Valley account for 4.2% of employment in non-IT STEM jobs. What creates such variance between San Francisco and another location, such as Indianapolis, and between STEM and IT occupations? Answering these questions entail disentangling the contributions of three potentially interdependent causes of the returns to skill – i.e., variance in the returns to different occupations, variance in returns across locations, and trends over time that influence locations and occupations to move in the same or different directions.

In this study, we analyze the 142 largest U.S. areas, known as core-based statistical area (CBSA). We focus on the period after the millennium, between the years 2000 and 2018 – a period of rapid change in IT that includes two recessions – near the early part of our data, and in the middle of it. The study compares the annual wage distribution for IT occupations with all STEM occupations (excluding IT), using the Occupational Employment Statistics (OES). Available statistics include the mean, median, top and bottom quartiles and deciles of the wage distribution. The analysis examines the determinants of two economic forces: factors shaping regional wage inequality, and the difference (between IT and STEM) in wage premium within the highest-paying cities.

The study begins by establishing two underappreciated stylized facts, which together frame a conundrum. First, the two decades after the turn of the millennium witnessed slower but more broad-based wage growth for IT labor than the experience during the dot-com boom. More to the point, the data show that IT wage growth after the millennium followed broader STEM trends. Locations with higher IT wages also have higher overall returns to STEM labor. The second stylized fact seems to be inconsistent with the first one: after the millennium, and especially after 2012, overall IT wage inequality increased substantially within the United States.

How can they be reconciled? The answer lies in the patterns of regional wage inequality. After the millennium, there is a growing wage gap between leading U.S. regions and other areas. A few local labor markets enjoyed a persistent wage premium relative to other areas, which increased by about 50% between 2000 and 2018. Our findings suggest the experiences in these locations – i.e., Santa Clara (Silicon) Valley, San Francisco, Seattle, New York, and Washington DC – account for almost the entirety of the increase in regional wage disparity in IT labor in the United States.

The next part of the paper characterizes the changes in regional wage inequality. We establish evidence of a pattern that runs against the notion that exceptionally high returns to the most valuable IT skills determine the wage premium in leading urban clusters. We find that the spread between top-decile and median wages negatively correlates with the average wage level in local IT labor markets and this negative correlation strengthened especially after 2012. Contrary to what most superstar models would predict, the top-earning IT workers were not disproportionately compensated in the leading cities. In the areas with thriving IT labor markets, all IT workers, and not just the top earners, benefited from the local wage premium. From the perspective of median-ability IT workers, moving to a tech cluster (e.g., San Francisco) reduces their wage difference from top-earning IT professionals. This pattern is surprising and not recognized by previous studies.

That insight frames the last part of the paper, analyzing the determinants of wage premium in top locations, and explaining the rising regional inequality in returns to IT labor. We collect data on regional determinants of labor market outcomes and examine how their impact on the regional difference changes over time. Using an Oaxaca-Blinder decomposition that compares the top five cities with other U.S. regions, we test among three broad categories of explanations for differences in IT and STEM wages. We label these categories as "capital complementarity," "agglomeration economies and labor pooling," and "innovation and entrepreneurship." We provide a summary of the findings as follows.

A leading explanation forecasts that regional wage premiums arise from complementarity between frontier IT assets and IT-skilled labor. Sectors that produce IT goods (including computers, electronics, and telecommunications) or use IT as input (such as finance, publishing, and business services) may offer exceptionally high wages to IT-skilled workers. IT-using and IT-producing firms choosing to cluster in the local area can raise IT labor market returns, sharing the benefits of matching complementary investments in IT assets and IT labor skills. These effects can be separately identified due to differences in the geographic spread of IT usage and production. We find, surprisingly, little empirical evidence supporting this explanation.

Another leading explanation stresses agglomeration economies among all skilled labor, where IT and STEM occupations are subject to the same economic factors. According to this explanation, agglomeration of professionals can lead to productivity spillovers among skilled labor, and high-quality matching between firms in need of skills and potential hires in high-density urban settings. The matching does not value the IT skills per sé but, instead, matches on facets of skilled labor, such as adaptability of employees to new requirements, sound judgment with discretion, and the ability to learn and perform cognitive tasks quickly. Empirically, we test for this explanation and find evidence consistent with it. For example, the share of highly educated labor explains a substantial fraction of the wage difference between top CBSAs and other U.S. cities. The population size of the area also accounts for a significant fraction of the regional wage gap.

The third explanation, which we label "innovation and entrepreneurship," also has received considerable attention. The IT sector evolves rapidly and has a long history of introducing novel ideas and business models. Many high-growth startups contribute to new job creation and productivity growth, seemingly boosting IT wages. Innovation and startup activities tend to cluster in a few urban locations, potentially contributing to the local wage premium. We do find evidence partially consistent with this explanation. Innovation (measured in patents) explains an increasing share of local IT wage premium in top-wage cities, especially after 2012. We do not find any support that entrepreneurship plays a role. Again, both results are surprising and contrast with the IT-producers' self-image as offering exceptional returns. Either factor is not a critical driver of high returns to IT labor – they are of moderate economic importance at best.

To summarize, we show that two distinct and competing forces shaped inequality trends in the IT labor market, over the last two decades. On the one hand, the advantage of tech hubs and urban metropolises – especially the combination of dense population and vigorous innovation – increasingly leads to higher IT wages, making some regions more attractive to skilled talent. On the other hand, wage spread narrowed within such advantaged areas, moving the top decile of IT wages into convergence with other STEM occupations. These findings are consistent with the view that provincial experiences within local labor markets shape the returns to skill bias and plays an increasingly significant role in driving wage inequality.

1.1 Related Work

This paper contributes to several strands of literature. First, it speaks to the literature documenting that advanced IT adoption led to substantial wage growth, concentrated in a small number of urban locations with dense population, a large supply of skilled labor, and high IT intensity. Advanced IT investments benefit wages in these areas disproportionately, exacerbating regional wage inequality between 1995 and 2000 (Forman, Goldfarb and Greenstein, 2002, 2005, 2012). Our findings contrast with this literature about the dot-com boom. We show that the relative growth in wages after the dot-com boom has been much milder but more broadly shared across regions, and more similar between IT and STEM (excluding IT) labor. Our contribution supports the view that the divergence of IT wages from other skilled wages in the 1990s was an aberration from long term trends for how skill-biased technical change shaped wages.

Second, this paper relates to the recent literature on new trends in the high-skilled labor market in the United States. Recent empirical evidence suggests that the theory of skill-biased technical change alone does not entirely explain the changes in the wage structure in the United States after 2000 (Goldin, Katz et al., 2020). Several papers offer explanations for patterns in U.S. wage inequality, including Deming (2017), Eckert, Ganapati and Walsh (2019), and Kaltenberg (2020), all broadly pointing to digital-enabled communication and falling costs for coordination across space as the mechanism driving changes in the wage structure of the U.S. skilled labor force. Consistent with this work, we also find that since 2012, the supply of college-educated labor explained a much smaller fraction of the overall top city wage premium. Instead, the wage structure appears to explain the increasing premium in the returns to IT skills in tech hubs and large metropolises. Our contribution also emphasizes the insights gleaned from comparing how IT wage patterns converge with and differ from similar skill types, i.e., STEM skills.

Third, this paper speaks to the literature on innovation spillovers, which are more geographically concentrated than typical industrial activity (Carlino and Kerr, 2015; Kerr and Robert-Nicoud, 2019; Rubinton, 2019; Moretti, 2019). New progress in the technology frontier creates transformative entrepreneurship, which may largely contribute to job creation and productivity growth (Decker et al., 2014). We find a moderate but statistically significant role of innovation in driving top-city advantages in providing higher wages to IT-skilled labor. The surprising finding is that this factor matters most in recent years and not before the 2010s.

Fourth, the paper adds to the large literature on the "great divergence" in wages across regions, according to which urban agglomeration contributes to the widening skilled wage gap (Moretti, 2013; Baum-Snow and Pavan, 2013; Baum-Snow, Freedman and Pavan, 2018). Like the literature, we find a large-city premium in IT wages. We also find that agglomeration appears to play a more significant role in recent years. Population size accounted for an increasing fraction of the local IT wage premium in top cities since 2012. Our results contribute new insights by analyzing the entire wage distribution for skilled labor and comparing across occupations across areas.

2 Background and Data

We shed light on the wage distribution of IT-skilled workers in the United States by analyzing patterns in regional IT labor markets. We focus on the comparison between IT and other STEM occupations, using aggregate annual wage statistics within each occupation category and geographic area (CBSA). This enables us to match it with data on characteristics of the regional labor market, as well as the demographic profile and industrial composition of major urban areas. This section describes these data sources and the other issues related to processing the data.

2.1 Aggregate Wage Statistics by Occupation and CBSA

We collect statistics on annual wage distribution across U.S. labor markets for all occupations and locations from 2000 to 2018. The data is available through the Occupational Employment Statistics (OES). The level of granularity for location is a core-based statistical area (CBSA). For occupation, detailed categories are defined by 6-digit codes under the Standard Occupation Classification (SOC). Wages are denominated in contemporary dollars (USD). Available statistics include the mean, median, top and bottom quartiles and deciles of the wage distribution.[1]

There are about 900 occupation codes in the United States, each defined by a 6-digit SOC code. The Standard Occupation Classification system makes adjustments to the occupational categories once every few years. The SOC codes to classify occupations changed in 2010 and 2017, and we use a crosswalk file to match these codes across years. We focus on a panel of annual wages in all IT occupations and other STEM occupations from 2000 to 2018.

According to the BLS Occupational Outlook Handbook, the Information Technology (IT) class consists of a few different categories, and the exact number of categories changed over time. All IT occupations belong to the broader class of around 90 STEM occupations. For the skill requirement of these occupations, jobs in most IT categories (except for IT support specialists) require at least a college degree. Some IT occupations have a higher average level of education, e.g., computer research scientists

[1] The public data may not contain all quantiles, e.g., in an area with a small population, and top wages are coded as missing. We impute missing quantiles by multiplying the next (lower) available quantile with the average ratio between the two quantiles in the overall data.

earned a master's or higher degree on average.

In 2010, the SOC system defined new occupations, replacing some of the old categories and expanding the number of IT occupations from 10 to 14.[2] We combine some of these categories into one group, to match the definitions before and after 2010. The resulting grouping of IT occupations (standardized across all years) consists of eight categories: Computer Programmers, Research Computer Scientists, Applications Software Engineers, Systems Software Engineers, Database Administrators, Computer Network Occupations, Computer Support Specialists, and Other Computer Occupations.

We assemble wage data in the largest metropolitan areas in the United States in terms of population size (in 2010). Our analyses, therefore, do not cover rural areas and small cities. We do not include small regions for several reasons. For a small region, wage data is more likely to be missing in at least one year. That typically arises due to few workers in a particular occupation, and either the statistics protect privacy by not publishing any data, or the region may not be sampled at all. Considerable evidence also

[2] After 2010, the occupation codes and names associated with IT include the following: 15-1111 Computer and Information Research Scientists; 15-1122 Information Security Analysts; 15-1121 Computer Systems Analysts; 15-1122 Information Security Analysts; 15-1131 Computer Programmers; 15-1132 Software Developers, Applications; 15-1133 Software Developers, Systems Software; 15-1134 Web Developers; 15-1141 Database Administrators; 15-1142 Network and Computer Systems Administrators; 15-1143 Computer Network Architects; 15-1151 Computer User Support Specialists; 15-1152 Computer Network Support Specialists; 15-1199 All Other Computer Occupations. The 2000 SOC system classified IT occupations into ten categories: 15-1011 Computer and Information Scientists, Research; 15-1021 Computer Programmers; 15-1031 Computer Software Engineers, Applications; 15-1032 Computer Software Engineers, Systems Software; 15-1041 Computer Support Specialists; 15-1051 Computer Systems Analysts; 15-1061 Database Administrators; 15-1071 Network and Computer Systems Administrators; 15-1081 Network Systems and Data Communications Analysts; 15-1099 All Other Computer Specialists.

suggests that some smaller regions and low density areas of the U.S. did not have access to frontier IT infrastructure during this time (Greenstein, 2020). Though this variance in supply and availability has been documented at sporadic intervals, for our purposes we would require consistent data about supply constraints across locations for close to two decades, which is not available to our knowledge.

The final data sample consists of annual wage statistics in 142 CBSAs spanning two decades. These statistics include moments of the wage distribution other than the mean. To illustrate the contents of the dataset, Table 1 includes the average wage for six IT occupations and for a representative sample of a dozen different STEM occupations. We show median and ninetieth quantile for San Francisco, which is near the high end of the wage distribution in 2018, Indianapolis, which is close to the middle, and Little Rock, which is near the lowest end. We also show their comparable wage levels for 2000, 2006, and 2012, and the comparative rank of these areas among the 142 areas.

Table 1 illustrates a few salient features of the dataset. First, the spread in wages across occupations is substantial. The IT occupation with the highest median wage in San Francisco is a computer and information research scientist, which is 90% higher than a computer support specialist, the IT occupation with the lowest median wage. A similar spread arises in other locations and in each year. Second, STEM wages deserve a similar observation. Among the highest-paying STEM occupations in San Francisco is an architectural and engineering manager, and it is 55% higher than a median STEM occupation (statistician), and 156% higher than the lowest STEM occupation (surveying and mapping technician). Again, a similar spread arises in other locations and in each year. Third, the spread between the highest and lowest location is quite substantial. Focusing just on computer and information research scientists (the highest paid IT occupation), San Francisco is 45% higher on average than Indianapolis, the location with the median, and the lowest-paying locations do not employ research scientists. Lastly, the rank of a location in the distribution appears to be persistent, but not fixed. It changes over time. While San Francisco remains either the highest or next highest location for IT wages, Indianapolis was ranked 81 in 2012, 58 in 2006, and 95 in 2000. Among the lowest in 2018, Little Rock was ranked, respectively, 128, 122, and 129. A goal of this study is to characterize that variance and persistence.

The data allows us to derive inequality measures and look into other properties of the wage distribution than the average wage level. Appendix A.2 and A.3 describe two methods to approximate the full distribution of annual wages. Both methods impose additional assumptions to fit the parameters that best approximate the full wage distribution within a given occupation category and CBSA.

3 Stylized Facts

This section presents a few stylized facts that demonstrate trends over the past two decades in returns to IT-skilled labor. We focus within and between region patterns and compare the trends in IT with those in STEM more broadly.

3.1 IT and Other STEM Wage Distributions over Time

STEM occupations are a significant share of the U.S. high-skilled labor market. As noted, among STEM occupations, IT occupations are a substantial category accounting for 40-50% of all STEM jobs. The total number of IT employees increased from about 3.4 million in 2000 to over 4.6 million in 2016. While IT occupations are perceived to be among the highest-paying professions even at the entry-level, it is an open question whether the past few years have witnessed rising returns to IT skills in comparison to other skilled labor.

In this section, we document stylized facts that together do not support the hypothesis that the growth in IT wage is "exceptional". We compare the wage distribution of IT labor with the rest of STEM labor. To do this, we use the aggregate wage statistics to approximate the entire wage distribution within an occupation in a given location. We simulate individual-level wage data from the approximate distribution and generate random samples of worker wages using total employment within a category as frequency weight. Appendix A.2 explains the details of this procedure to construct random samples based on aggregate wage moments.

The first observation is that the distribution of IT wages and that of STEM wages appear broadly similar from 2000 to 2018. Figure 1 shows the annual wage distributions of IT wages (solid line) and other STEM wages (dashed line) spanning the last two decades. The density functions of the returns in the two types of labor markets have similar means and spreads and they do not show different trends over time. This evidence is inconsistent with the hypothesis that IT jobs provide particularly high returns setting it apart from the rest of the skilled labor force.

There is another side to the same coin. Across various occupations, STEM wages have increased steadily between 2000 and 2018. The average STEM worker earned about 50,000 USD in 2000, and 80,000 USD in 2018.[3] The average wages appear to grow at similar rates in the IT labor market as in the market for other types of STEM labor.

The IT wage distribution has become increasingly unequal since the 2000s. In Figure 1, the shape of the density function flattened and the mass at the tails of the wage distribution increased over time. That pattern masks the underlying causes. It still could be the case that the increasing influence of tech companies in Silicon Valley and other large cities may have led to booming local IT job opportunities that benefit only a relatively small part of the workforce. It also could be the case that technological progress has increased the returns to top-skilled workers, though encountered other factors that depressed it at the same time. This motivates our later investigation into the causes of regional variance.

3.2 Rising Regional Inequality in IT Wages

In this subsection, we examine the second dimension of wages, regional variance. This analysis provides evidence that regional inequality accounts for a large part of the rise in overall wage inequality in IT occupations. We find that the IT sector opportunities are concentrated in a small number of places, which led to rising wage inequality between regions. Booming markets in some cities benefit workers across the local wage distribution,

[3] Wages are measured in contemporary dollars.

possibly at the expense of shrinking opportunities in other locations and out-migration from declining areas.

To assess the extent to which regional factors contribute to overall inequality in the IT labor market, we decompose total wage inequality into within-region and between-region components. The Occupational Employment Statistics (OES) data contains only aggregate statistics, such as the mean, median, top and bottom quartiles, and deciles of the wage distribution within an occupation category and a CBSA. Appendix A.3 describes how we approximate the wage distributions using parametric assumptions and estimate the parameters to fit the aggregate statistics. The decomposition into between- and within-region components can be directly calculated from the estimated parameters, after weighting by employment shares across regions.

Figure 2 plots a between-region inequality measure for IT and other STEM wages annually from 2000 to 2018. The Y-axis of Figure 2 is an inequality index – the Generalized Entropy index GE(1), also referred to as Theil's T. The left panel shows the between-region Theil's T in all 142 largest CBSAs in the United States, among IT wages (solid line) and other STEM wages (dashed line), respectively. The right panel shows the between-region Theil's T, after excluding the top CBSAs that pay the highest IT wages among all U.S. regions. The excluded CBSAs are a few tech clusters and superstar cities – Silicon Valley, San Francisco, Seattle, Washington DC, and New York City.

Patterns in Figure 2 suggests between-region inequality increased from 2000 to 2018 in both IT and other STEM labor markets, but particularly sharply for IT occupations[4]. While the regional disparity in IT wages grew by 60% since the early 2000s, wages of non-IT STEM labor increased only by about 30%. The trajectories of regional wage disparity between IT and non-IT STEM labor diverged particularly after 2012, as the wage difference between top cities and the rest of the country increased sharply in IT but not other parts of the STEM labor market.

The rise in regional inequality in IT labor markets can be attributed to a small number of CBSAs with the highest overall returns to IT-skilled workers. The five CBSAs that appear to contribute to the entire increase in regional inequality are regions that have historically been the most attractive places for IT talent. These regions are also where most high-growth IT companies locate their headquarters. After excluding these regions from the decomposition, between-region inequality appears to be flat over time, in both IT and non-IT STEM wages. In the rest of large U.S. urban areas, the STEM labor markets experience overall similar returns in the 2000s as in the 2010s.

In Figure 3, we plot the distribution of IT wages in the top five CBSAs (solid line) and the rest of large metropolitan areas (dashed line), every six years over a two-decade period from 2000 to 2018. The wage distribution in top CBSAs diverged from the distribution in other large cities. It shifted to the right gradually over time, indicating that the average wage pulled ahead of the rest of the large CBSAs in the country.

To summarize, IT wage growth has shown increasingly similar patterns

[4]Different measures of inequality decomposition print broadly similar pictures of the time trajectory of between-region wage inequality. Appendix A.3 presents the same decomposition using alternative inequality decomposition approaches, including Theil's L and total variance. Theil's L is more sensitive to differences at the bottom of the wage distribution relative to Theil's T.

to wage growth in the STEM labor market, but it is also distinct in several ways. A small number of superstar cities have been pulling ahead of the rest of the country in returns to IT-skilled labor, while their advantages in STEM wages have been lower and remained relatively stable since the early 2000s.

4 Regional Wage Premium and Inequality

This section examines whether regional premium in IT wages are driven by extremely high returns to the top-skilled IT workers, what the literature has labeled as superstar effect. In recent years, frontier IT skills are demanded only by large firms in only a few places such as Silicon Valley and New York City, but these skills are highly rewarded and pay much more relative to older and more traditional IT skills (Tambe, 2014).

To assess this question empirically, we compare the average level and the 90th-50th percentile spread of the wage distribution within each CBSA. If the top end of the wage distribution drives the widening regional gap between local IT labor markets, then we should expect to see strong correlation between the wage spread in a CBSA and the average wage level in that location. On the other hand, if the local labor market rewards the average-skilled worker, then the wage spread would not positively correlate with wage level within a CBSA.

This approach is less effective when data on top wages are missing, which is the case for some data in the Occupational Employment Statistics (OES). OES masks some wage statistics if they are above a threshold that varies from year to year. The top quantiles in CBSAs with higher returns are more likely to be missing. We impute missing wage quantiles using the ratio between moments of the wage distribution in other parts of the data.

In addition, the wage data does not contain information about non-wage compensation, and therefore may not fully reflect the overall returns to IT work. For example, some jobs with high earning potential (e.g., founders of high-growth IT startups) reward workers with stock options, which are not recorded in the OES data. While this is unlikely to affect the middle of the wage distribution in the vast majority of the occupations we examine, it can have consequences for creating new recipients in the upper tail of entrepreneurially-oriented IT and STEM occupations, the so-called 1% of income, particularly during prosperous economic years. This is an important open question.

4.1 Decomposition of Wage Level and Spread

To focus on the location-specific component of wage levels and spreads, we need to account for compositional differences that drive overall wage levels of IT workers. For example, if firms in Location A employ more programmers than location B, where more research scientists receive employment, then we need to account for whether programmer positions or research scientists receive higher compensation (typically, the latter do). The total difference in wages between these locations will partially reflect the difference in the composition of IT jobs in various occupations, rather than the location-specific premium for all IT jobs within the local labor market.

To account for differences in the composition of occupation within each location, we estimate the following fixed-effects regression models in Equations 1 and 2. These regression models control for occupation-specific fixed effects, so that estimates of location-specific fixed effects β_{jt} and γ_{jt} reflect solely the wage premium associated with advantages of any given location, rather than the composition of jobs with intrinsically different characteristics and returns. We call these estimated fixed effects "indexes" of wage levels and spreads.

$$\log(W_{ijt}^{avg}) = \beta_{jt} + \xi_{it} + \epsilon_{ijt} \tag{1}$$

$$\log\left(\frac{W_{ijt}^{p90}}{W_{ijt}^{p50}}\right) = \gamma_{jt} + \psi_{it} + \nu_{ijt} \tag{2}$$

In the equations above, W_{ijt}^{s} denotes the annual wage statistic in occupation i, CBSA j, and year t. The wage statistic denoted by s takes values from among the set of moments in the available data, including the average, the median, and the top and bottom quartiles and deciles.

The estimation is conducted for each year separately, on annual wage statistics from 2000 to 2018. To estimate the wage level index, we use only the average wage of each location and occupation, in the outcome variable of equation 1; to derive the wage spread index, we use the ratio between the top decile and median wage, in the outcome variable of equation 2.[5]

We focus on wage differences at the higher end of the wage distribution (instead of, e.g., the median to bottom decile ratio) for the wage spread index, because our goal is to test whether top-skilled IT

[5]The fixed-effects estimation in equation 1 identifies the difference in the effects between each location (and similarly, occupation and statistic type) and a base CBSA. We need to omit one occupation and one CBSA each to estimate the contribution of all other occupations and CBSAs relative to the base category. To be consistent across years, we choose the same base categories in all years of regressions. We select San Francisco (CBSA code 41860) as the omitted location, and Computer Programmer (2010 SOC code 15-1021) as the omitted occupation.

workers drive average returns to IT skills to be particularly high in top-wage regions.

Appendix Table A.1 and A.2 list CBSAs with the largest wage level index and wage spread index respectively. Each column shows the average index over a 3-year period: 2000-2002, 2006-2008, 2011-2013, and 2016-2018. The omitted category (San Francisco) has an index of 0 by default, and therefore all the other indexes are relative to San Francisco. For example, a location j with a wage level index of β_{jt} pays on average $\left(\beta_{jt} \times 100\%\right)$ higher in IT wages than San Francisco in year t.

To relate the level and spread of the wage distribution within CBSAs, we estimate the regression model in Equation 3. We estimate the equation for IT workers and other STEM workers separately.

$$\hat{\gamma}_{jt} = \sum_{\tau=2000}^{2018} \alpha_\tau \left[\beta_{jt} \times \mathbb{1}(t=\tau)\right] + e_{jt} \tag{3}$$

Each coefficient α_τ for a particular year τ estimates the correlation between the wage level index $\beta_{j\tau}$ and the wage spread index $\gamma_{j\tau}$ of CBSA j in year τ. A negative α_τ indicates that in a location with a higher average wage, the within-CBSA inequality is generally lower among IT (or non-IT STEM) workers.

4.2 Correlation Between Regional Wage Level and Spread

This subsection describes the results of estimation, which shows the trends in location-specific wage indexes over time and present coefficient estimates of the relationship between the level and spread indexes in each CBSA. Appendix A.4 contains a partial list of CBSAs with the highest estimated level and spread indexes.

The patterns in regional wage differences persist over time. The highest-wage locations have been the same CBSAs from the early 2000s to present day. Tech clusters, especially Silicon Valley and San Francisco Bay Area have sustained substantial premiums in returns to IT-skilled labor over the last two decades, where IT jobs pay at least 10–15% more than in a city ranked the 6th in wage level index. The wage premiums in these locations have increased even more in recent years compared to the 2000s. For example, IT workers in San Francisco earn about 7% higher wages in the early 2000s relative to those in the 6th-ranked CBSA, but this SF wage premium increased to a much larger 13% in 2016 – 2018.

On the other hand, the lowest-earning locations are not that different from middle-earning places. IT jobs in CBSAs with the lowest wage level indexes pay about 50% less than in those in Silicon Valley and San Francisco, but the gap did not grow between 2000 and 2018.

The set of locations with the highest wage spread indexes consists of an entirely different list of CBSAs, compared to places where the wage level premiums are the highest. It is not "the usual suspects," and we believe there is information in pattern because none of the top-wage locations make it to the list of "most unequal" CBSAs, which have the largest top-decile to median ratios in within-region IT wage distribution. More broadly, wage levels and spreads are negatively correlated across locations, particularly for IT occupations after 2012. Figure 4 plots the relationship between the wage spread index and the wage level index, in 2000, 2006, 2012, and 2018 respectively. The relationship between these indexes for the IT labor market slopes downward since 2012. Importantly, this latter pattern that does not generalize to other STEM labor markets.

Figure 5 plot all the coefficients from estimating Equation 3 in multiple years from 2000 to 2018. It shows the time trend in the relationship between wage levels and spreads within CBSAs, for IT occupations (solid line) and non-IT STEM occupations (dashed line) respectively. The coefficients for IT occupations were significantly below zero in most years after 2012, while it diverged away from STEM trends around the same time. This pattern appears persistent and does not seem to be associated with the 2008 financial crisis. It arises in years after the economy started to recover from the aftermath of the Great Recession.

Two main trends emerged from analyzing the differences in regional premium in local IT labor markets across large metropolitan areas in the United States. On the one hand, the gap between the top-earning CBSAs and the rest of the country widened gradually over the two decades after the millennium, a trend that notably accelerated in the most recent years. On the other hand, locations paying higher average levels of wages have become more attractive locations to median-skilled workers, and the tech-hub premiums in returns to IT skills are not limited to paying top-ability workers extremely high salaries. Most important, these trends are specific to regional IT labor markets – they are not observed broadly in other STEM labor markets outside of IT.

5 Oaxaca-Blinder Decomposition of IT-Wage Premium

Previous sections have revealed diverging regional trends as the dominant pattern in IT wages over the last two decades. The goal of the empirical exercise in the remainder of the paper is to show the extent to which different regional factors explain differences in IT wage premium. What factors contribute to the different labor market experience for IT occupations in the leading cities? We propose to use and modify the Oaxaca-Blinder decomposition to test the relevance of different explanations.

5.1 Comparing Across Locations

We conduct an Oaxaca-Blinder decomposition over the total IT wage difference between top CBSAs and other areas. This approach focuses on explaining the wage premium in the top five CBSAs relative to the rest of the U.S. labor market for IT skills. These top five CBSAs include San Jose-Sunnyvale-Santa Clara MSA (Silicon Valley), San Francisco-Oakland-Hayward MSA, Seattle-Tacoma-Bellevue MSA, New York-Newark-Jersey City MSA, and Washington-Arlington-Alexandria MSA. To test each explanation, we include one explanatory variable at a time. After this, we pick all the explanatory variables that explain a substantial portion of the group difference to add to the regressions. The last step estimates the relative contribution of each factor to the total wage gap between top CBSAs and the rest of the locations.

The first step in the Oaxaca-Blinder decomposition procedure involves estimating separate OLS regression models of wages on all the explanatory factors for each group. Then the same model is estimated on the pooled data containing both groups, resulting in reference coefficients that are taken to be parameters of an "unbiased" model assumed absent any group discrimination. The explained difference attributed to each regional factor is directly calculated from the difference between the coefficient estimates of the group-specific models and the reference coefficients.

Each factor X_{jt} is a CBSA-level variable for location j that can possibly explain the wage premium of top locations in year t. The first stage of Oaxaca-Blinder estimates the following OLS model in Equation 4, for the set of top CBSAs and the rest of the locations separately.

$$\log\left(W_{ijt}^{m}\right) = \beta X_{jt} + \gamma Controls_{jt} + \psi_{t}^{m} + \xi_{it} + \epsilon_{ijt}^{m} \tag{4}$$

Each observation is associated with an aggregate wage level statistic m, for each occupation i and CBSA j. The estimation weighs each observation by the size of employment. The model in Equation 4 controls for occupation fixed effects ξ_{it} to account for differences in occupational composition across regions. We also control for statistic fixed effects ψ_{t}^{m}. We use all available quantiles of the wage distribution, not only the average wage level, to estimate Equation 4.

One advantage of the decomposition is its ability to highlight changes in different explanations over time. Technological progress may increasingly reward some firms (e.g., large established corporations versus small young businesses) and particular tasks (e.g., cognitive non-routine jobs and communication tasks) more than others. Hence the explanatory power of different regional features may shift over time, as they interact with changes in organizational structures and job functions in IT-intensive industries. The other advantage is related. The decomposition provides explanations consistent with the descriptive facts in Sections 3 and 4 about changes in IT labor market inequality patterns over time. For example, the wage gap between top CBSAs and other locations has been rising consistently over the last two decades, but particularly after 2012, which marks somewhat diverging trends between IT and other STEM skills.

This approach also has one drawback. Changes in skill requirement and job tasks associated with an occupation are not reflected in the OES wage data. Such shifting within-occupation trends can directly affect returns to IT labor, which cannot be distinguished from explanations based solely on observed regional factors. Related, as the same occupation code may include a different set of tasks over time, the nature of jobs in each IT occupation may have evolved over two decades. The same occupation category may entail more frontier skills in some CBSAs than others, particularly given the geographic concentration of some frontier IT.

5.2 Regional Characteristics and Explanatory Variables

For the decomposition exercise we collect data on regional characteristics. For more details about data collection and variable construction for regional features, Appendix A.1 contains a thorough discussion. Several explanations provide paths to testing potential location-driven explanations for changes in IT wage inequality patterns over time. We organize data collection around these explanations.

IT Capital Complementarity and Marshallian Externalities

One set of explanations focus on the complementarity between the usage of IT capital and skilled labor. Firm-level IT intensity can affect labor market returns to complementary professional activity. Rising top-city premiums in IT wages also may be due to distinct demand-side characteristics of firms producing IT. Many of the most valuable public companies are in IT-intensive sectors, for example, and tend to be collocated. Many IT-producing firms also tend to be IT-intensive. Prior work has, fortunately, developed methods for distinguishing between high and low IT-intensity, and we follow these definitions (Jorgenson et al., 2005; Forman, Goldfarb and Greenstein, 2012), but with updated analysis to account for more recent changes in practice (Calvino et al., 2018). More specifically, we use the global indicator defined in Calvino et al. (2018) to classify as IT-using the set of sectors with medium-high or high digital intensity in both 2000s and 2010s. [6] We also use a traditional definition for IT-producing sectors and add communications to account for the spread of digitization.[7]

[6] The indicators are grouped by International Standard Industrial Classification (ISIC) Rev. 4 codes, which we map into 5-digit NAICS industries using a crosswalk file from Census Bureau.

[7] We classify industries as IT-producing and IT-using by the International Standard Industrial Classification (ISIC) Rev. 4 codes. The following sectors are IT-producing: Computer, electronic and optical products (26), Telecommunications (61), and IT and other information services (62-63). The following industries are IT-using: Wood and paper products, and printing (16-18), Computer, electronic and optical products (26), Electrical equipment (27), Machinery and equipment (28), Transport equipment (29-30), Furniture, other manufacturing, repairs of computers (31-33), Wholesale and retail trade, repair (45-47), Publishing, audiovisual and broadcasting (58-60), Telecommunications (61), IT and other information services (62-63), Finance and insurance (64-66), Legal and accounting activities (69-71), Scientific research and development (72), Advertising and market research, other business services (73-75), Administrative and support service activities (77-82), Public administration and defense (84), and Other service activities (94-

We next use County Business Patterns (CBP) data to measure the number of establishments in a given industry (NAICS) and CBSA. The CBP data is reported at an annual frequency between 1999 and 2017 and contains the number of establishments of different sizes within each CBSA and 6-digit NAICS industry. We construct the share of establishments in IT-using and IT-producing sectors within each location. Even though a firm may have multiple establishments across different CBSAs, we are interested in the share (rather than the total number) of IT-intensive firms, using establishment counts to construct the IT-intensity measures associated with a given location is a reasonable approach. It is also plausible that large firms differ in productivity levels from small firms, and contribute disproportionately to local labor market patterns. Therefore, we also use data on the number of establishments within different ranges of employment sizes to construct a measure for the share of large IT-intensive establishments (defined as those employing at least 500 workers) for IT-using and IT-producing sectors respectively.

Agglomeration Economies and Labor Pooling

Another set of explanations highlights the role of urban agglomeration and the size of the local labor market. At least three mechanisms lead to advantages of the urban environment in generating high-earning opportunities for skilled labor. First, a large labor market facilitates the matching between firms and the pool of local workers (Duranton and Kerr, 2015). Large urban areas increasingly draw inflows of migration of highly-educated workers (Diamond, 2016). When an abundant supply of

96).

high-skilled workers concentrates in the local area, this high-quality talent pool attracts more firms to co-locate with the local labor market, in order to gain better access to these highly productive workers as potential hires. According to this explanation, agglomeration helps matches on facets of skilled labor, such as adaptability to new requirements with rapid learning, sound judgment with discretion, and the ability to perform cognitive tasks quickly.

A second mechanism highlights that the demand for IT skills is exceptionally high in a small number of superstar cities. The geographic concentration of tech activity may create a shortage for particular types of frontier IT skills, bidding up IT wages in local areas where such skills are in high demand. IT workers possessing frontier skills (or the ability to acquire these skills quickly) are a scarce resource for firms, even in regions with relatively abundant overall labor supply. This contributes to an IT wage premium as firms align themselves in digital transformation across a variety of industries.

Finally, the concentration of highly skilled workers in an urban location generates productivity spillovers among local workers, who may or may not possess similar skill. For example, such knowledge spillovers are associated with agglomeration economies that boost inventor productivity (Moretti, 2019), and innovation appears to benefit more from clustering in concentrated locations than industrial activity more broadly (Carlino and Kerr, 2015).

To test this set of explanations, we use CBSA-level demographic profiles to construct variables that measure the size of the local labor market and supply of educated workers. Specifically, we collect demographic statistics for each CBSA in the sample, using the American Community Survey data available from 2005 to 2017. We also use the 1999 census data to construct the same set of local demographics, but for each MSA in the year of 2000. Some of the variables include total population and share of college-educated adults (aged 25 and above) within each CBSA. We also construct a set of control variables related to regional characteristics from the American Community Survey (post-2005) and U.S. Census (2000) data. These include control variables, such as the age and racial composition of the population living in the area. They also include net migration inflows, as well as poverty and unemployment statistics.

Patents, Innovation, and Entrepreneurship

A third set of explanations stresses variance in regional innovation, high-growth entrepreneurship, and business dynamism. Startups and young firms are essential contributors to job creation and productivity growth. Before 2010, young firms generated about one-sixth of all new jobs and startups account for 20 percent of gross job creation but less than 10 percent of all firms (Decker et al., 2014). Although most startups fail, a small fraction of firms that survive become highly productive, and experience transformative growth and have a disproportionate impact on productivity and employment in the economy.

Invention tends to occur in geographically concentrated clusters, measured by aggregate venture capital and patenting activities (Kerr and Robert-Nicoud, 2019). Recent research documents the rising geographic concentration of IT patents in the United States (Forman and Goldfarb, 2020). In addition, evidence suggests innovation enhances firm productivity and raises worker compensation (Kline et al., 2019). The commercialization of innovation and frontier-stretching activities can lead to higher wages in occupations related to, but not directly engaged in invention activities. The IT sector is vibrant with startup entry and dynamic changes, where new launches of products and services frequently redefine boundaries of the industry and leading trends.

We proxy for inventive activity within a CBSA using the total number of Computers & Communication patents assigned to inventors in the CBSA (Forman and Goldfarb, 2020). To measure business dynamism and entrepreneurial activity, we use Kauffman Indicators (available on Kauffman Foundation website) and newly registered businesses (available from the business formation data by the Census Bureau). The data are both aggregated to the state level and available from 1999 to 2017. In Kauffman Indicators, "transformative" high-growth entrepreneurship is separately measured from other types of new firms (which may include small businesses that do not grow in size). The Kauffman Indicator is a z-score constructed from combining four variables that track startup activity in the U.S. across states annually. These four variables are the rate of new entrepreneurs, opportunity (non-subsistence) share of new entrepreneurs, startup job creation, and startup survival rate.

The patent data are available for each CBSA, while the Kauffman data are available only at the state level. In only a few large states (specifically, California and Texas), the Kauffman indicator are coarser than ideal because there are more than a few large CBSAs with varying labor market characteristics and business formation dynamics. However, most states have very few CBSAs that are large enough to make it into our IT wage data sample. Therefore, state-level indicators reflect variations in local startup activities in the largest CBSAs for most of our data.

5.3 Empirical Results and Discussion

We start by testing each factor by including them in the Oaxaca-Blinder decomposition, one variable at a time. Table 2 shows the list of explanatory factors we test, and whether they appear to explain a non-trivial portion of the wage difference between top cities and the rest of large metropolitan areas at some point over the last two decades. Other background regional characteristics are also added as control variables, including the age and racial composition, poverty status, unemployment rate, and migration inflow in each CBSA.

The measures associated with IT-capital complementarity appear to explain very little of the top-CBSA wage premium. The share of overall establishments in IT-intensive sectors (either IT-using or IT-producing) does contribute to the total wage gap. Large establishments (with 500 or more employees) in IT-using sectors appear to have a moderate association with the wage premium in top locations. It appears industrial clusters do not contribute much to the advantage of CBSAs in rewarding IT skills. This suggests that demand-side factors may have been less critical in shaping returns to IT labor than anticipated.

The factors associated with agglomeration spillovers and labor pooling appear to robustly predict the top CBSA premium in IT wages in recent years. The share of college-educated adults and the total population size explain a substantial part of the wage differences from 2016 to 2018.

The patterns around the contribution of innovation and entrepreneurship to top region wage premiums are a bit more complicated. They appear to have started to account for a moderate but significant part of the regional inequalities after 2013 and shortly after the economy began to recover from the Great Recession. However, in the 2000s and early 2010s, regional innovation activities did not explain in any non-trivial degree the advantages of top CBSAs in rewarding IT-skilled labor.

After identifying the regional factors that explain a non-trivial portion of the wage premium in top locations, we assess their relative contributions to regional IT wage inequality over time. Figure 6 visualizes the results of this assessment, showing the explained share of the total wage gap each year between 2006 and 2018 (left panel), along with the percentage contribution of each factor to total explained group difference (right panel).

Consistent with what we have found in the descriptive section, the wage premium in top cities have been rising over time, and increased particularly sharply after 2012. Figure 6 left panel also suggests that known factors identified from theory do not sufficiently explain the sharp increase in the wage gap post-2012. The explained portion of the wage gap remained stable around the same level, suggesting that the differential changes in wage structure between top CBSAs and the rest of the country may have been a primary reason for the higher top-location premium in IT wages after 2012.

One explanation of differential changes in "wage structure" between top CBSAs and other locations is that technical progress may have transformed digital business models, particularly among firms in the Silicon Valley and other tech clusters, but less so elsewhere.

The right panel of Figure 6 breaks down the contribution of each factor to the share of top-CBSA wage premium explained by differences in observed characteristics. All results are scaled to reflect the fraction explained by each variable. After 2012, the fraction of wage difference explained by the local supply of high-skilled labor fell by almost 40%. The fraction of wage difference explained by population size also decreased from 2012 to 2013 but increased year after year since then.

Interestingly, the fraction of wage difference explained by IT-related patents increased after 2012, and account for 6% of the total explained IT wage gap between top CBSAs and other locations. This number is moderate compared to the relative contribution of skill supply, but it demonstrates trends of increasing relevance over time, especially since 2012.

These results suggest that agglomeration economics and innovation in information technology may have contributed increasingly to the regional gap in IT wages in recent years. The trends reflect the wage premium in the top clusters and largest metropolises more precisely, rather than the general difference between, e.g., the 10th ranked city and the bottom city in terms of IT wages. We focus on explaining this particular type of regional gap in IT wages because most of the changes in IT labor market inequality over the last two decades reflect a few top cities pulling ahead of the rest of the country, as documented in Section 3.

Table 3 shows Oaxaca-Blinder decomposition results at different points in time over the last two decades. Column 1 shows the results estimated from wage statistics in 2000. Columns 2 – 4 show results on pooled samples over three-year periods starting in 2006, 2011, and 2016. The regional variables in 2000 come from the 1999 census, and those in all later years come from the American Community Survey, which did not start to collect data until 2005. Hence the regional factors and control variables are not available between 2001 and 2004, for which we cannot assess each factor's relative contribution.

There is a substantial IT wage premium associated with a larger pool of high-skilled (college-educated) labor in all years. This suggests that the supply of skills in the local labor market has been a primary driver of productivity and IT wages throughout the last two decades.

The coefficients on population and IT-related patents in Table 3 have increased over time, suggesting that agglomeration spillovers and technical innovations have played a more significant role in driving top CBSA wage premium in recent years. On the other hand, population size was not a driver of regional advantage in 2000, where the coefficient on the logarithm of population size was negative. Neither did IT patents contribute much to top-location wage premium in 2000.

Between 2016 and 2018, the top CBSAs paid about 25.5% more in IT wages than other large U.S. metropolitan areas. Observed regional factors, including demographics and other control variables, explained 69% of this wage gap. While the high-skilled share (college-educated adults) accounted for 25% of the explained difference, population size accounted for 10.2%, and IT-related patents 6.3% of the explained difference.

The increasing importance of technological innovation in driving IT wage premiums in top cities may be explained by frontier IT skills (e.g., big data), which tend to be sought by the largest firms in a small number of tech clusters (Tambe, 2014). Returns to frontier skills have been growing quickly, relative to other parts of the IT labor.

Other control variables, especially the share of the white population and population below the poverty line, appear to explain a moderate fraction of top-location premium in IT wages since 2006. The relative explanatory power of these demographic control variables did not vary systematically over the last decade.

The IT wage structure in the United States appears to experience changes in recent years that affect top cities differently from the rest of the country. The "unexplained" share of overall wage premium in top CBSAs has been growing and accounted for almost the entire increase in the regional wage gap since 2012. The same set of local factors from previous years cannot fully explain this more recent increase in top-city wage premium.

6 Conclusion

This study examines whether exceptional factors shape returns to IT-skilled labor rather than being part of general STEM wage trends. We also analyze changes over time in the contribution of different factors in regional labor markets to the IT wage premium.

The evidence points to rejecting any narrative that stresses IT-exceptionalism after the dot-com era, and that suggests the rise of IT wages in that era may have been an aberration from long term trends. We conclude that IT wages increasingly followed the patterns consistent with long-run skill-biased technical change. Returns for highly skilled IT workers are not substantially higher than those to the rest of the IT labor market. Instead of idiosyncratic IT wage trends, general economy-wide forces associated with skill-biased technical change appear to shape the overall wage structure of STEM labor more broadly. Since the end of the dot-com boom in the early 2000s, wages of IT-skilled labor and STEM labor in fields other than IT have evolved onto similar paths.

Large urban areas feature an IT wage premium above and beyond overall returns to STEM labor. Such an advantage concentrates in only a few superstar cities – Silicon Valley, San Francisco, Seattle, New York, and Washington DC – which pulled ahead of all other U.S. regions in terms of IT wage and employment. Shifts in the composition of occupations in local labor markets do not shape these patterns. Instead, they reflect wage growth within occupational categories.

While the popular sentiment may surmise that exceptionally high returns to IT skills contribute to rising wage inequality within the United States, we document a surprising and previously unrecognized effect working in the opposite direction. We find empirical support for the declining spread between the top and median IT wage earners, in some urban areas and particularly locations with higher average IT wages. The correlation between earnings average and spread within a location has been significantly negative in recent years.

Since the recovery from the Great Recession, IT wages have become less unequal in top CBSAs than elsewhere, at the same time as these top cities experience faster productivity and wage growth. Other U.S. metropolitan areas appear to be lagging behind, and place-based policies to stimulate productivity growth and improve job opportunities may be particularly useful to these areas.

Using an Oaxaca-Blinder decomposition approach, we analyze the underlying causes behind the spread between cities. We find that an increasing amount of top-CBSA wage premium is not explained by regional factors that previously explained a large part of the regional wage gap before 2012. Top-CBSA wage premium grew particularly rapidly after 2012 and most of this growth is explained by differential changes in the IT wage structure in top cities relative to the rest of the country.

The supply of high-skilled labor appears to have accounted for a substantial portion of around 16% of total IT wage difference between the top CBSAs and other regions. One source of the advantage of star cities is their superior talent pool of highly skilled workers. As firms adapt to rapid digital transformation, high-quality IT labor is an increasingly valuable resource that experiences rising demand even in locations with relatively abundant skill supply.

Agglomeration economies and innovation appear to play an increasing role in driving the advantage of tech clusters and large metropolises in attracting IT talent. The explanatory coefficients on population size and IT-related patents have increased since 2012. On the other hand, aggregate IT-intensity of local firms did not contribute substantially to rising top-city wage premiums in recent years.

These findings have implications for economic policies. Our results also suggest that a range of regional growth policies to recruit IT producers, which have been commonly pursued over the last few decades, have not paid off with higher local wages. Instead, recruitment of any professionally skilled and innovative labor led to distinctively high wage growth. Policy focused on the IT wage premium or on digital skills would have been too narrow an approach. Instead, a more effective policy should encourage the development of STEM-based skills more generally. In addition, while innovation may have contributed to IT worker productivity, policies to encourage innovation and entrepreneurship had, at best, only a moderate effect on wage growth. Among the most effective policies were those that made a location attractive to skilled workers and their employers.

Our results raise numerous questions about the interplay between regional composition of labor supply and the demand for distinct types of skill, which creates the regional variance in skill-biased technical change. Our findings reinforce models that stress the benefits to scale that operate at the regional level. Yet, that also leaves some questions open. For example, what type of competition between a few superstar cities generates a negative correlation between mean levels and spread of wage premiums? Our findings also raise numerous questions about models of superstar cities that solely depend on the size of cities, or the matching of supply and demand of specialized labor. Competition between cities must play a role in shaping the composition of skill-biased returns within the cities.

Finally, our findings used the ups and downs of macroeconomic trends as markers for change in patterns over time. While our sample was limited to two decades of wage premiums, we implicitly adopted the view taken throughout the literature that these processes operate over long horizons and decadal time scales. That too raises questions about how the opportunities and limits associated with a growing national economy or stagnated economic activity shapes regional comparative advantages and the effect of skill-bias on wages.

References

Autor, David H, Frank Levy and Richard J Murnane. 2003. "The skill content of recent technological change: An empirical exploration." *The Quarterly journal of economics* 118(4):1279–1333.

Autor, David H, Lawrence F Katz and Alan B Krueger. 1998. "Computing inequality: have computers changed the labor market?" *The Quarterly journal of economics* 113(4):1169–1213.

Autor, David H, Lawrence F Katz and Melissa S Kearney. 2006. "The polarization of the US labor market." *The American economic review* 96(2):189–194.

Autor, David et al. 2019. *Work of the Past, Work of the Future*. National Bureau of Economic Research.

Baum-Snow, Nathaniel, Matthew Freedman and Ronni Pavan. 2018. "Why has urban inequality increased?" *American Economic Journal: Applied Economics* 10(4):1–42.

Baum-Snow, Nathaniel and Ronni Pavan. 2013. "Inequality and city size." *Review of Economics and Statistics* 95(5):1535–1548.

Byrne, David and Carol Corrado. 2019. Accounting for Innovation in Consumer Digital Services: IT Still Matters. Technical report National Bureau of Economic Research.

Calvino, Flavio, Chiara Criscuolo, Luca Marcolin and Mariagrazia Squicciarini. 2018. "A taxonomy of digital intensive sectors.".

Carlino, Gerald and William R Kerr. 2015. Agglomeration and innovation. In *Handbook of regional and urban economics*. Vol. 5 Elsevier pp. 349–404.

Decker, Ryan, John Haltiwanger, Ron Jarmin and Javier Miranda. 2014. “The role of entrepreneurship in US job creation and economic dynamism.” *Journal of Economic Perspectives* 28(3):3–24.

Deming, David J. 2017. “The growing importance of social skills in the labor market.” *The Quarterly Journal of Economics* 132(4):1593–1640.

Diamond, Rebecca. 2016. “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980-2000.” *American Economic Review* 106(3):479–524.

Duranton, Gilles and William R Kerr. 2015. The logic of agglomeration. Technical report National Bureau of Economic Research.

Eckert, Fabian, Sharat Ganapati and Conor Walsh. 2019. “Skilled tradable services: The transformation of US high-skill labor markets.” *Available at SSRN 3439118* .

Forman, Chris and Avi Goldfarb. 2020. Concentration and Agglomeration of IT Innovation: Evidence from Patenting. In *The Role of Innovation and Entrepreneurship in Economic Growth*. University of Chicago Press.

Forman, Chris, Avi Goldfarb and Shane Greenstein. 2002. Digital dispersion: An industrial and geographic census of commerical internet use. Technical report National Bureau of Economic Research.

Forman, Chris, Avi Goldfarb and Shane Greenstein. 2005. “How did location affect adoption of the commercial Internet? Global village vs. urban leadership.” *Journal of urban Economics* 58(3):389–420.

Forman, Chris, Avi Goldfarb and Shane Greenstein. 2012. "The Internet and local wages: A puzzle." *American Economic Review* 102(1):556–75.

Goldin, Claudia and Lawrence F Katz. 2007. The race between education and technology: The evolution of US educational wage differentials, 1890 to 2005. Technical report National Bureau of Economic Research.

Goldin, Claudia, Lawrence F Katz et al. 2020. Extending the Race between Education and Technology. Technical report National Bureau of Economic Research.

Greenstein, Shane. 2020. Digital Infrastructure. In *Economics of Infrastructure Investment*. University of Chicago Press.

Horton, John J and Prasanna Tambe. 2019. The Death of a Technical Skill. Technical report Working Paper.

Jorgenson, Dale W, Mun S Ho, Kevin J Stiroh et al. 2005. "productivity, Volume 3: information technology and the American growth Resurgence." *MIT Press Books* 3.

Kaltenberg, Mary. 2020. The Knowledger Manager: Wage Premiums in Knowledge Diverse Industries. Technical report.

Kerr, William and Frédéric Robert-Nicoud. 2019. "Tech Clusters.".

Kline, Patrick, Neviana Petkova, Heidi Williams and Owen Zidar. 2019. "Who profits from patents? rent-sharing at innovative firms." *The Quarterly Journal of Economics* 134(3):1343–1404.

Leibbrandt, Murray, Arden Finn and Ingrid Woolard. 2012. "Describing and decomposing post-apartheid income inequality in South Africa." *Development Southern Africa* 29(1):19–34.

Moretti, Enrico. 2013. "Real wage inequality." *American Economic Journal: Applied Economics* 5(1):65–103.

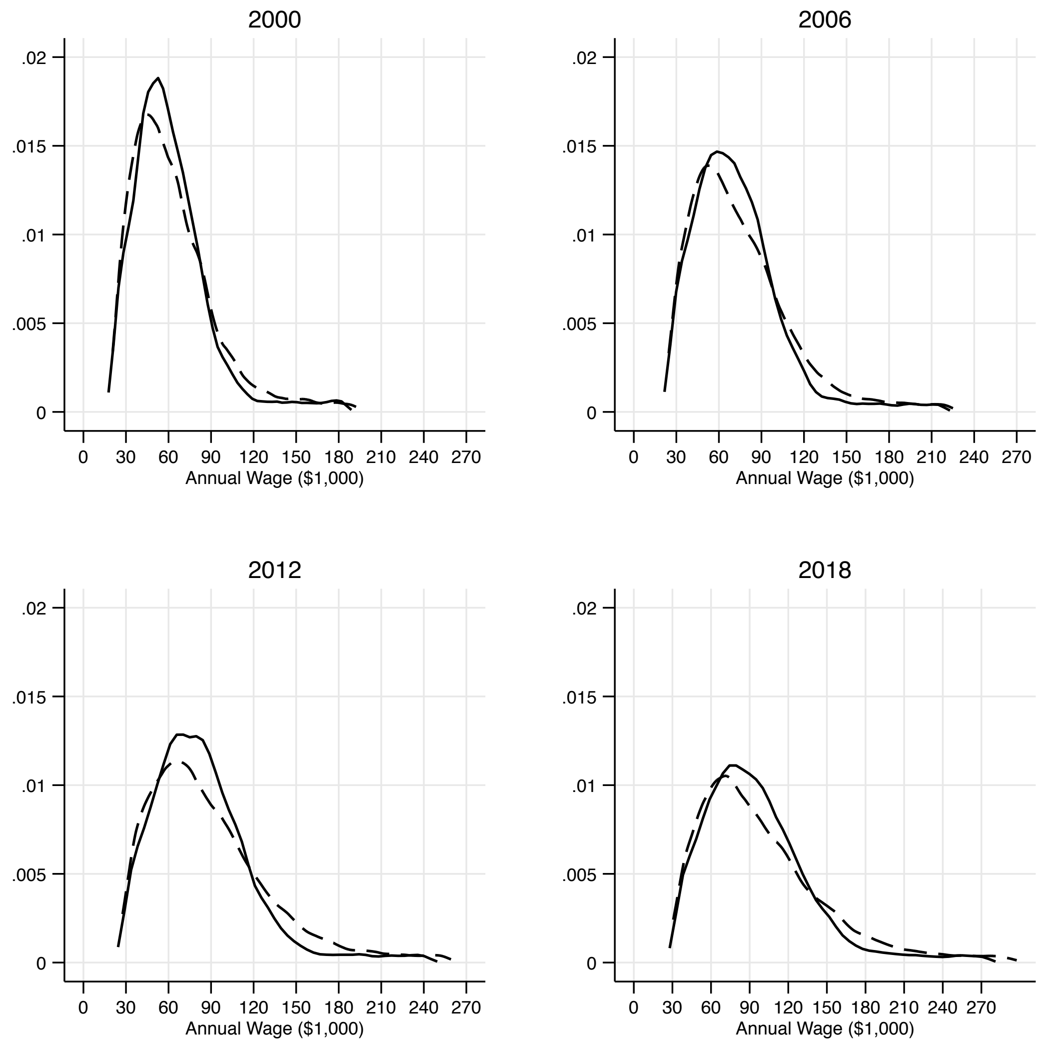
Moretti, Enrico. 2019. The effect of high-tech clusters on the productivity of top inventors. Technical report National Bureau of Economic Research.

Rubinton, Hannah. 2019. "The Geography of Business Dynamism and Skill-Biased Technical Change.".

Tambe, Prasanna. 2014. "Big data investment, skills, and firm value." *Management Science* 60(6):1452–1469.

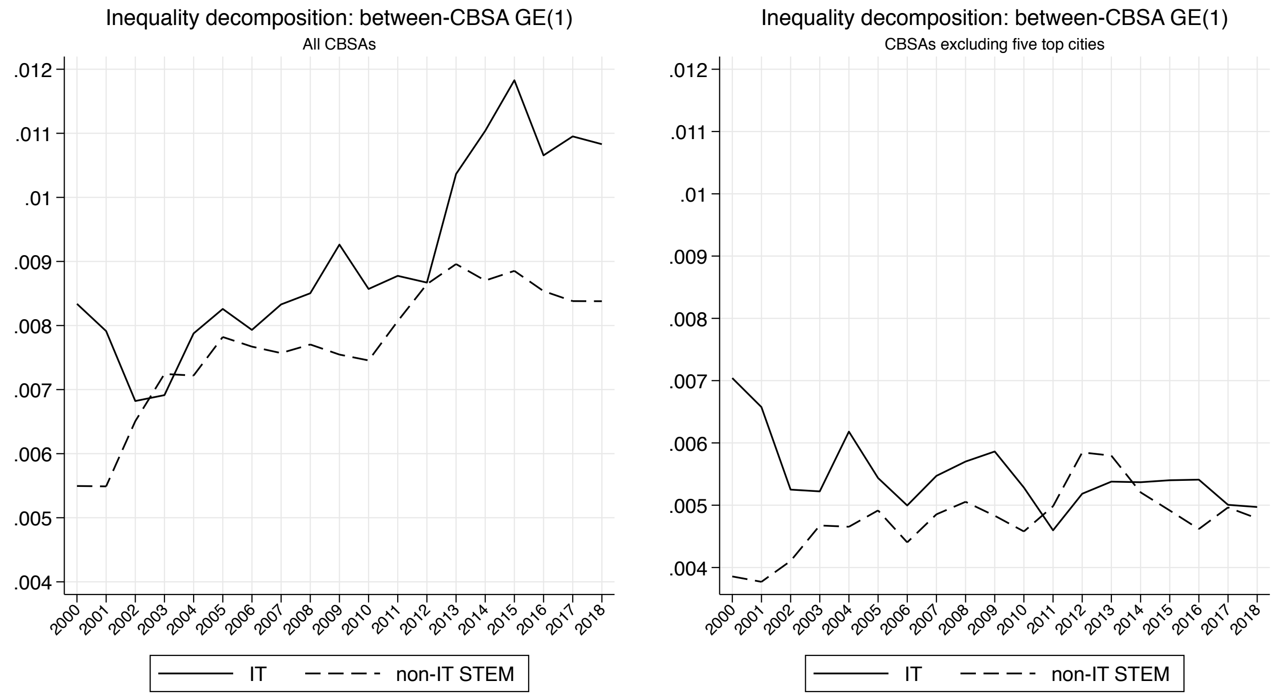
Tambe, Prasanna and Lorin M Hitt. 2012. "The productivity of information technology investments: New evidence from IT labor data." *Information Systems Research* 23(3-part-1):599–617.

Figure 1: Overall IT and Other STEM Wage Distributions



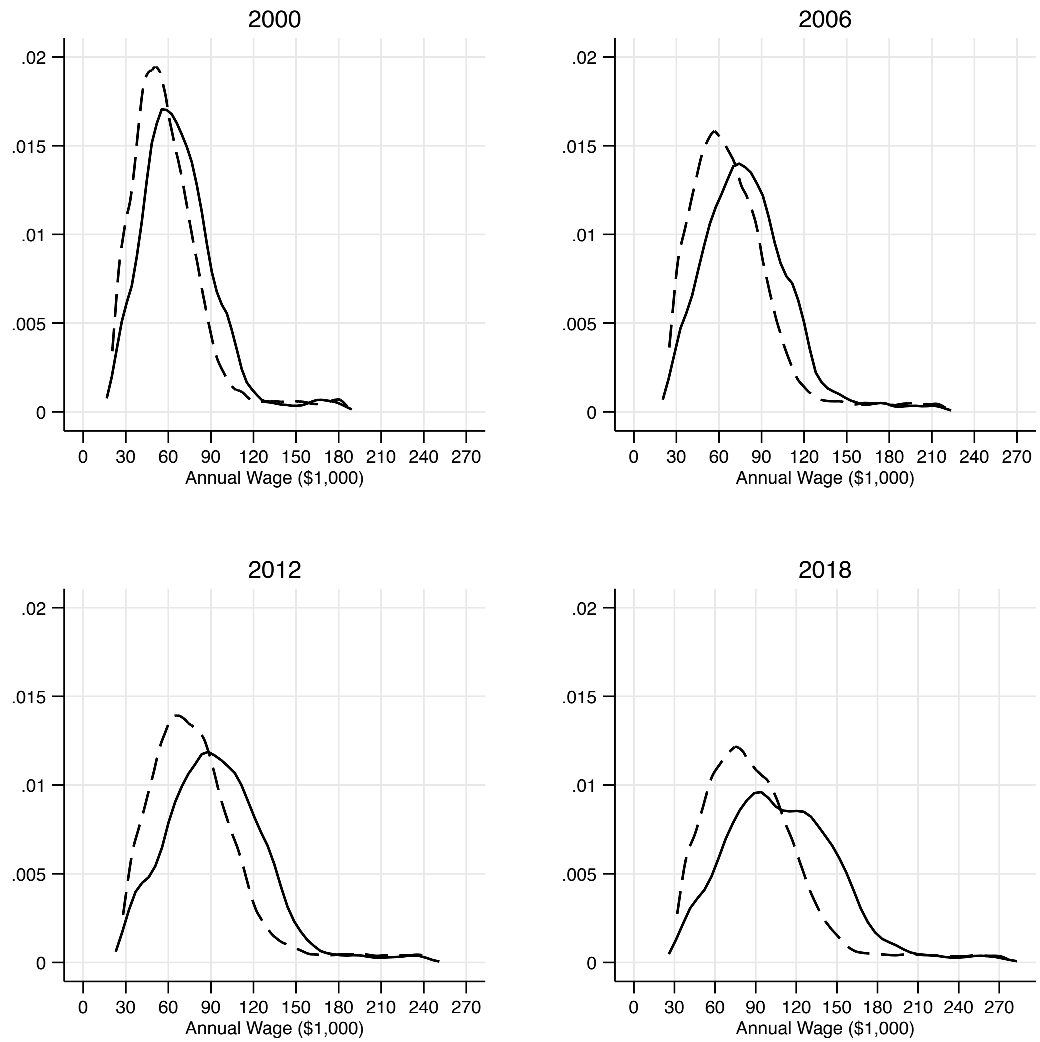
Notes: This figure shows snapshots of wage distributions (in contemporary dollars) for STEM occupations, in four years 2000, 2006, 2012, and 2018. The solid lines represent the estimated kernel density function of wages in IT occupations, while the dashed lines represent the estimated kernel density function of wages in non-IT STEM occupations in all large U.S. cities. The X-axes shows wage earned annually (in thousand USD), and all graphs are fixed to be comparable on the same scale and axis ranges.

Figure 2: IT and Other STEM Wage Inequality Between CBSAs



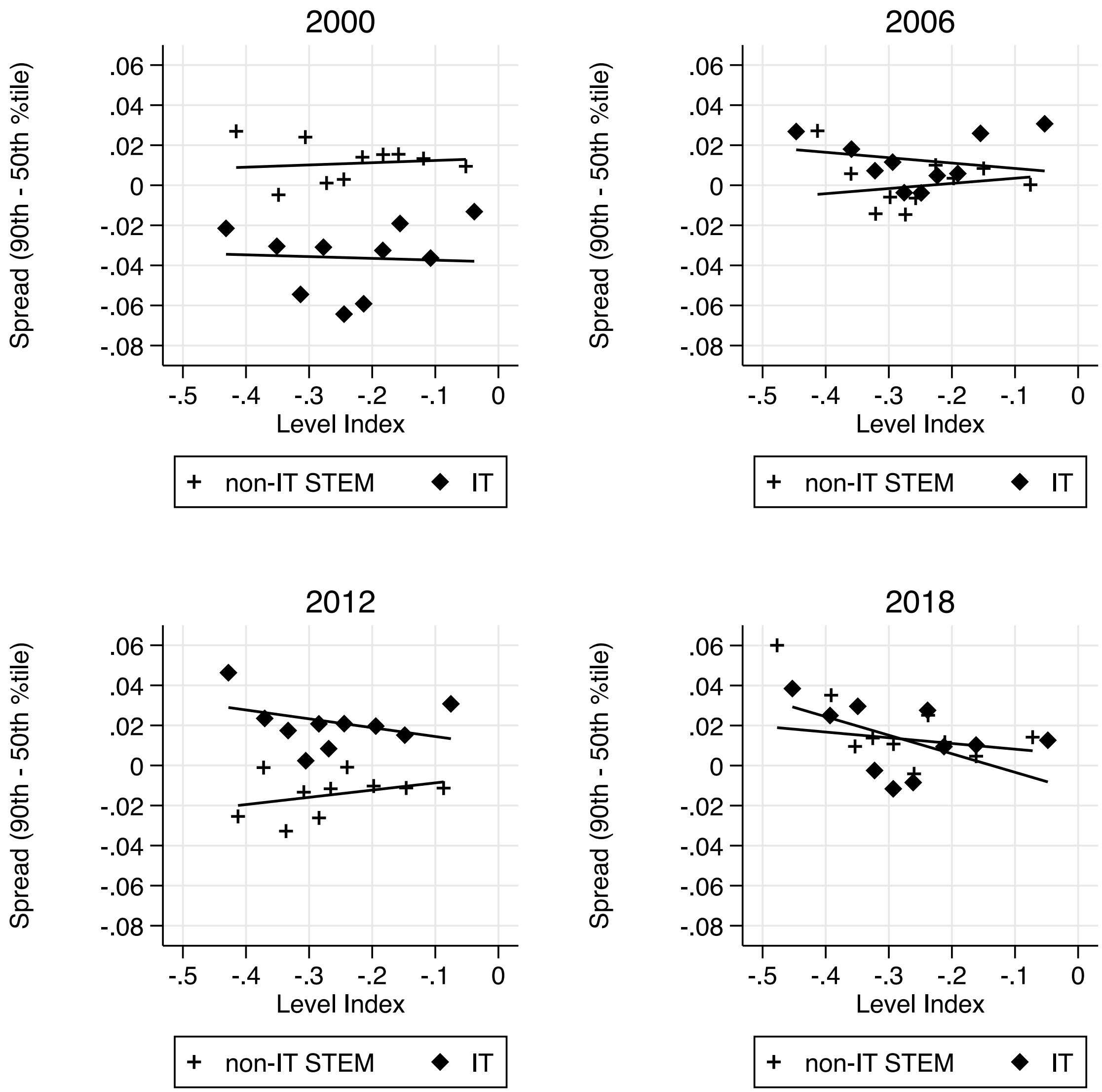
Notes: This figure shows the between-CBSA wage inequality among IT occupations (solid) and non-IT STEM occupations (dashed) respectively, in each year between 2000 and 2018. The X-axes represent each year from 2000 to 2018. The Y-axes represent the between-CBSA component of an inequality index. The inequality index is measured using Theil's T – General Entropy Index GE(1), which is more sensitive to the top part of the wage distribution, and constructed from CBSA-occupation level wage statistics (mean, 10th, 25th, 50th, 75th, and 90th percentiles). The left panel plots annual regional wage inequality for all CBSAs, and the right panel plots the same measure but excluding five areas – Silicon Valley, San Francisco, Seattle, New York City, and Washington DC.

Figure 3: Difference in IT Wage Distribution Between Top Locations and Other CBSAs



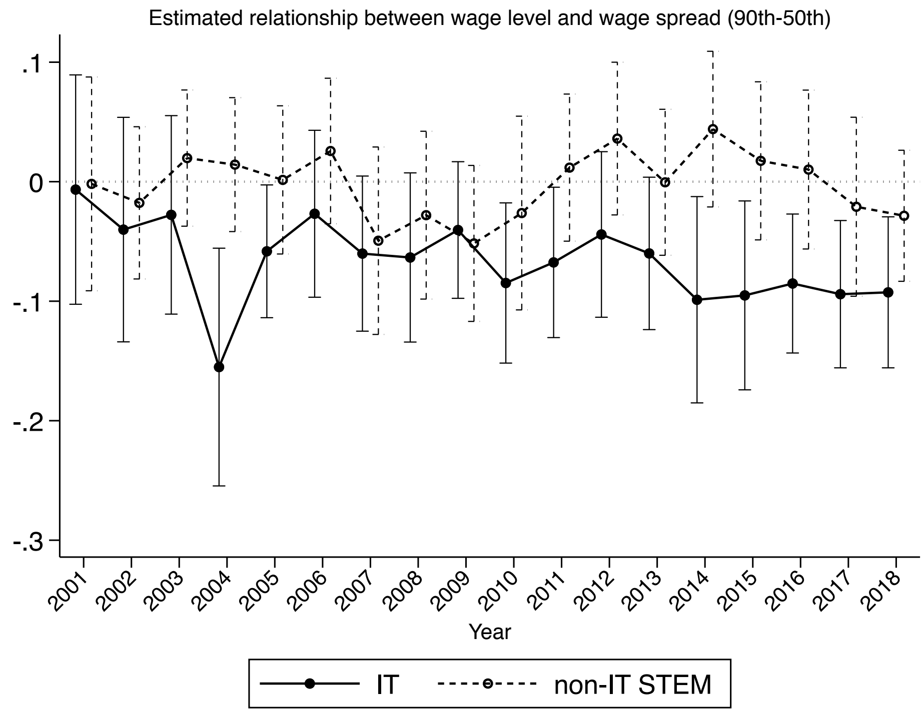
Notes: This figure shows snapshots of wage distributions (in contemporary dollars) for IT occupations, in four years 2000, 2006, 2012, and 2018. The solid lines represent the estimated kernel density function of IT wages in the top CBSAs – Santa Clara (Silicon) Valley, San Francisco, Seattle, New York City, and Washington DC. The dashed lines represent the estimated kernel density function of IT wages in all other large U.S. cities. The X-axes shows wage earned annually (in thousand USD), and all graphs are fixed to be comparable on the same scale and axis ranges.

Figure 4: Binscatter Plots: Wage Level and Spread Indexes



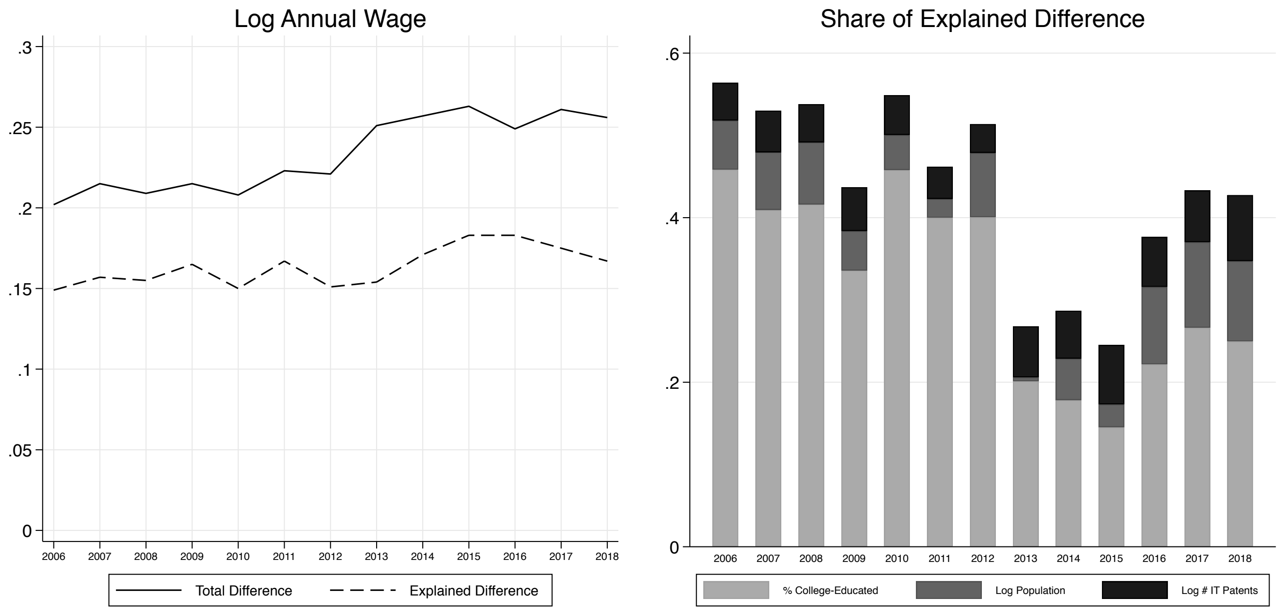
Notes: This figure plots the relationship of CBSA indexes between wage level and wage spread, for IT occupations (diamond shape) and non-IT STEM occupations (plus mark) respectively in each year of 2000, 2006, 2012, and 2018. The X-axes represent the wage level index – the CBSA fixed-effect estimates from the wage level decomposition regressions. The Y-axes represent the 90th-50th percentiles wage spread index – the CBSA fixed-effect estimates from the wage spread decomposition regressions.

Figure 5: Relationship Between Wage Level and Spread Indexes, All Years



Notes: This figure plots the relationship of CBSA indexes between wage level and wage spread, for IT occupations (solid) and non-IT STEM occupations (dashed) respectively from 2000 to 2018. The X-axis represent the year, and the Y-axis represent the correlation between the 90th-50th percentiles spread and wage level indexes, estimated from the decomposition regressions.

Figure 6: Contribution to Top City IT-Wage Premium of Explanatory Factors Over Time



Notes: This figure shows the total wage premium in top cities, and the contribution of major explanatory factors to the wage difference across locations. The results are derived from an Oaxaca-Blinder decomposition on annual wages. The left panel shows, from 2006 to 2018, the total difference (solid) and explained difference (dashed) between average wages in top locations – Santa Clara (Silicon) Valley, San Francisco, Seattle, New York City, and Washington DC – and the rest of large CBSAs in the United States. The right panel shows the relative contribution, of each regional explanatory factor, to total explained wage difference, from 2006 to 2018.

Table 1: Annual Wage Summary Statistics by Occupation and CBSA

Annual Wages ($ Thous.)	**San Francisco, CA**			**Indianapolis, IN**			**Little Rock, AR**		
	Mean	P50	P90	Mean	P50	P90	Mean	P50	P90
IT Occupations (2018)									
Computer Scientists, Research	141	137	216	117	112	165	/	/	/
Software Engineers, Systems	137	135	193	87	82	123	83	79	118
Software Engineers, Applications	144	138	206	89	82	128	85	82	126
Database Administrators	108	109	160	79	77	119	79	78	113
Computer Programmers	106	106	151	85	76	130	76	72	115
Computer Network Occupations	112	110	161	83	78	124	70	66	102
Computer Support Specialists	74	71	109	50	47	76	47	46	64
Average Level and Ranking by Year	Mean	Rank		Mean	Rank		Mean	Rank	
2000	69.5	2		53.2	55		43.4	124	
2006	84.8	2		63.0	51		52.6	122	
2012	98.6	2		72.8	53		61.8	120	
2018	120.9	2		79.7	71		70.0	121	
Non-IT STEM Occupations (2018)									
Architectural and Engineering Managers	184	170	259	122	118	171	118	118	165
Aerospace Engineers	133	130	171	97	94	130	120	120	167
Statisticians	119	117	169	71	60	122	59	52	93
Architects	100	96	157	79	72	129	73	68	115
Mechanical Drafters	72	71	110	52	49	75	51	53	75
Surveying and Mapping Technicians	72	72	108	43	40	62	39	37	57
Average Level and Ranking by Year	Mean	Rank		Mean	Rank		Mean	Rank	
2000	71.7	2		52.0	95		47.9	129	
2006	86.4	3		67.4	58		57.7	122	
2012	106.1	2		76.3	81		66.7	128	
2018	121.0	2		87.9	65		75.7	127	

Notes: This table shows the average wage moments in the final data sample. We select three locations, at the top, middle, and bottom parts of IT wage level in 2018 across all 142 major CBSAs. We list the mean, median, and top decile of 2018 wages in each of these locations for selected STEM occupations. We also list the aggregate level and ranking (among the 142 CBSAs) of each location in 2000, 2006, 2012, and 2018, in terms of wages in all IT and non-IT STEM occupations respectively.

Table 2: List of Regional Explanatory Factors to Top City IT-Wage Premium

Explanatory Factors	2016 - 2018
Agglomeration spillovers	
Share college-educated	**17.7%**
Total population	**12.5%**
IT capital complementarity	
# IT-producing establishments	0.1%
# large IT-producing establishments	-0.1%
# IT-using establishments	-1.0%
# large IT-using establishments	**1.8%**
Innovation and entrepreneurship	
Number of IT-related patents	**6.4%**
Kauffman Entrepreneurship Index	1.7%

Notes: This table reports the results from testing regional factors potentially explaining the average wage difference between top locations – Santa Clara (Silicon) Valley, San Francisco, Seattle, New York City, and Washington DC – and the rest of the large CBSAs. For three categories of explanatory theories, the table shows the percentage explained by each regional factor between 2016 and 2018.

Table 3: Oaxaca-Blinder Decomposition Results

	Top-City Premium in Log (Annual Wages)			
	2000	2006 - 2008	2011 - 2013	2016 - 2018
	(1)	(2)	(3)	(4)
Total Difference	0.175***	0.209***	0.232***	0.255***
	(5.08)	(11.08)	(11.57)	(11.38)
Agglomeration Economies and Labor Pooling				
Share of College-Educated Adults	0.037***	0.064***	0.052***	0.042***
	(4.42)	(21.48)	(14.67)	(13.09)
Log (Total Population)	-0.009*	0.011***	0.006**	0.018***
	(-2.51)	(6.62)	(3.11)	(10.10)
IT Capital Complementarity and Marshallian Externalities				
Log (# Large IT-Using Establishments)	-0.000	0.002***	0.002***	0.000
	(-0.04)	(3.93)	(4.28)	(0.11)
Patents, Innovation, and Entrepreneurship				
Log (Total IT Patents)	0.005	0.007***	0.007***	0.011***
	(0.97)	(4.56)	(5.28)	(8.05)
Control Variables				
20 <= Age < 35	0.005	0.008***	-0.003**	-0.002*
	(1.37)	(8.87)	(-3.28)	(-2.11)
35 <= Age < 45	0.049***	0.006**	-0.005**	-0.003
	(7.17)	(2.59)	(-2.61)	(-1.43)
45 <= Age < 55	0.010*	-0.002	0.011***	0.007***
	(2.52)	(-1.84)	(9.64)	(4.90)
55 <= Age < 65	-0.000	0.003*	-0.000	-0.003***
	(-0.19)	(2.06)	(-0.66)	(-5.72)
Age >= 65	-0.013***	0.000	0.001*	0.006***
	(-4.11)	(1.10)	(1.97)	(6.88)
Unemployment Rate	-0.003	-0.010***	-0.007***	-0.006***
	(-1.87)	(-11.65)	(-7.66)	(-8.52)
Share Below Poverty Line	0.002	0.020***	0.022***	0.018***
	(0.73)	(13.79)	(9.48)	(9.67)
White	0.078***	0.052***	0.098***	0.040***
	(6.52)	(11.49)	(16.48)	(5.94)
Black	-0.009**	-0.009***	-0.005*	0.002
	(-2.70)	(-5.57)	(-2.11)	(1.96)
Native American	0.000	0.001***	0.001***	0.001***
	(0.35)	(3.60)	(4.24)	(4.87)
Asian	-0.074***	-0.036***	-0.063***	0.001
	(-5.29)	(-7.16)	(-8.93)	(0.09)
Net Population Inflow	0.030	/	-0.002*	-0.006***
	(1.68)	/	(-1.98)	(-3.39)
Missing Population Inflow	-0.020	/	0.002	0.004*
	(-1.42)	/	(1.59)	(2.57)
Unexplained Difference	0.053*	0.055***	0.074***	0.080***
	(2.11)	(4.16)	(5.17)	(4.43)
No. Obs.	4914	15342	15570	15648

Notes: This table presents coefficient estimates from the Oaxaca-Blinder decomposition on regional differences in wages, between the top locations – Santa Clara (Silicon) Valley, San Francisco, Seattle, New York City, and Washington DC – and the rest of large CBSAs in the United States. Each column shows decomposition results that include the average wage of each of the two groups being compared, as well as the total difference, divided into an explained (endowment) and an unexplained (structural) component. It also lists the contribution of each explanatory regional factor to total difference. Column 1 shows results on wage data in 2000. Columns 2 – 4 show results on wage data over three-year periods, starting in 2006, 2011, and 2016, respectively. The estimation of all columns use aggregate data at the CBSA-occupation level, weighting observations by total employment, and pool together all available statistics, including the mean, the 10th, 25th, 50th, 75th, and 90th percentiles of the subgroup's wage distribution.

A Appendix

A.1 Data Sources for Regional Explanatory and Control Variables

To assess explanations to regional variation in wages, we also collect data on features of each major metropolitan area. These data come from public sources, and the following discussion summarizes data sources and key variables.

American Community Survey. We collect data on regional features by CBSA. We use the American Community Survey data on land area size, population and population density, and demographic profile (e.g. education attainment, unemployment rate, age composition, etc). Data availability starts in 2005.

Census 1999. The Census of 1999 provides demographics by metropolitan statistical areas. The variables available are similar to those in the American Community Survey, but for an earlier year. We obtain information about MSA-level population, median household income, education level, and ethnicity.

County Business Patterns. We use data from County Business Patterns to measure regional concentration of a number of industries, e.g. ICT (information and communication technology), FIRE (finance, insurance, and real estate), manufacturing, healthcare, etc. Measures include employment size, annual payroll, number of establishments, and number of large (over 100 workers) establishments. Data availability starts in 1999.

USPTO Patent Data. We measure the number of patents by category in each MSA, to capture the data to measure the aggregate number of ICT-related patents in each MSA over time, to proxy for the amount of IT-related innovation. Data is available for all years from 1999 to 2017.

Business Formation and Entrepreneurship. For new business formation and entrepreneurial activities, we use two sets of aggregate data – Census Business Formation and Kauffman Indicators[8]. These data are available at the state level, for all years from 1999 to 2017.

The frequency of all the data are at the annual level. Regional explanatory and control variables except business formation and Kauffman indicators are available at the level of core-based statistical area, and available for all 142 major metropolitan areas which have the largest population in the United States. We use regional features as regressors with a 1-year lag. We do this to account for the fact that some of the changes in the surroundings may take time to lead to changes in labor market outcomes.

A.2 Approximating Full Wage Distributions from Aggregate Statistics

The kernel density function for wage distributions in Section 3 is estimated on a simulated random sample, rather than an actual representative sample of STEM workers across large metropolitan cities in the United States. To generate such a random sample involves a few steps and some assumptions, as this section will describe.

The BLS Occupational Employment Statistics (OES) data contains, for

[8] Data available from https://www.census.gov/programs-surveys/bfs.html and https://www.kauffman.org/historical-kauffman-index/microdata respectively.

each occupation (defined by the Standard Occupation Classification codes) and CBSA, the average and key quantiles of the distribution of annual wages. The key quantiles include the 10th, 25th, 50th, 75th, and 90th percentile wages. Data are masked without being reported for higher quantiles, if the value exceeds a certain threshold which varies across years. Therefore, we have a right-censored raw data set, which renders the data set not directly comparable across years consider this issue and a number of other differences in the underlying sampling scheme to generate these statistics.

We approximate the cumulative density function of the wage distribution based on these summary statistics. To do this, we use linear interpolation on the wage quantiles, and make an assumption about the minimum of the wage numbers that is reasonable given the log-normal shape of the wage distribution, and then derive the maximum wage to fit the average wage level. After we get the approximated CDF of each distribution, we use inverse transform sampling to generate a random sample that from the CDF.

Each CBSA and occupation has a different wage distribution, and employ different numbers of workers. To create a full random sample, we sample from the cumulative density function for each occupation and CBSA, using the total employment in each group as weights. The scaling factor is 100 – for each 100 workers in a particular category, we generate one data point randomly from the interpolated density function. The full random sample consists around 1.1 million simulated data points on STEM workers across the 150 largest MSAs from 2000 to 2018. Around 47% of these simulated observations are IT worker wages.

A.3 Decomposing Wage Inequality into Between- and Within-CBSA Components

We use standard measures of wage inequality, such as Theil's T (GE1), to decompose total wage inequality into a within-region component, and a between-region component. Other inequality measures suitable for the decomposition are Theil's L (GE0) and variance. For data generated according to a log normal distribution, such decomposition entails closed-form formulas that are easy to compute.

We do not have individual-level wage data for a representative sample of workers, hence cannot decompose the wages without imposing assumptions on the shape of the wage distributions. Instead, we have a number of summary statistics on wages (i.e., the mean, 10th, 25th, 50th, 75th, and 90th percentiles) in each occupation and CBSA (for the largest 150 metropolitan area in the United States). The summary statistics of occupational and regional wages suggest high similarity between the raw wage distribution and the family of log normal distributions. Therefore, we approximate the wages with a log normal distribution within each aggregate group, and calculate the scale and shape parameters that best fit the available wage statistics, by minimizing the sum of squared deviations.

We fit a log-normal distribution to the log of the quantile and mean statistics in the data, to minimize the sum-of-squared-errors. Given the quantiles, and assume functional form for the CDF of the wage distribution, assuming Y is within $(0, \infty)$. Let $\log Y \sim N(\mu, \sigma^2)$.

$$Pr\left(Y \leq y\right) = \frac{1}{2} + \frac{1}{2} erf\left[\frac{\ln y - \mu}{\sqrt{2\sigma^2}}\right]$$

Then the theoretical mean of Y is

$$E[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{5}$$

Fitting the quantiles gives for each $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$

$$y_\tau = \exp\left(\mu + \sqrt{2\sigma^2} \cdot erf^{-1}(2\tau - 1)\right) \tag{6}$$

We have at most 6 equations and 2 unknown (μ, σ^2). We can solve for μ and σ^2 by minimizing the sum of least-squares deviations of the log of these statistics

$$\mathbb{L}(\mu, \sigma) = \left(\mu + \frac{\sigma^2}{2} - \log \bar{y}\right)^2 + \sum_\tau \left(\mu + \sqrt{2}\sigma erf^{-1}(2\tau - 1) - \log y_\tau\right)^2$$

Taking the first-order conditions of Equations 5 and 6, we get the following system of equations

$$\begin{aligned}
(N+1)\mu + \frac{\sigma^2}{2} - \log \bar{y} + \left[\sqrt{2}\sum_\tau erf^{-1}(2\tau - 1)\right]\sigma &= \\
\sum_\tau \log y_\tau \\
\left(\mu + \frac{\sigma^2}{2} - \log \bar{y}\right)\sigma + \left[\sqrt{2}\sum_\tau erf^{-1}(2\tau - 1)\right]\mu + 2\sum_\tau \left[erf^{-1}(2\tau - 1)\right]^2 \sigma &= \\
\sqrt{2}\sum_\tau erf^{-1}(2\tau - 1) \log y_\tau
\end{aligned}$$

Solving the above system of equations gives estimates of μ and σ^2 that parametrizes the wage distribution in the log-normal family to minimize the total least-squared errors in fitting the available wage statistics.

Using the estimated parameters, we can then derive approximations to inequality measures for each category, as well as aggregate these measures into an overall within-region component and a between-region component.

We focus on the class of Generalized Entropy measures of inequality, e.g., Theil's T (GE1) and Theil's L (GE0). The theoretical values of these measures for a given random variable Y are calculated as

$$\begin{aligned} GE(0) &= \ln E[Y] - E[\ln Y] \\ GE(1) &= \frac{E[Y \ln Y]}{E[Y]} - \ln E[Y] \end{aligned}$$

If we restrict Y to have a log normal distribution with scale and shape parameters (μ, σ^2), then the expressions for the inequality measures can be simplified, because both $GE0$ and $GE1$ are equal to $\frac{\sigma^2}{2}$ under the assumption that Y is log-normal.

This simplification makes regional decomposition particularly straight-forward. The goal is to use category-level estimated wage distribution parameters to derive aggregate inequality measures as well as divide these measures into between-region and within-region components.

We construct a number of inequality measures (i.e total variance and generalized entropy) and decompose them into within- and between-region components according to the standard procedures in the literature, such as described in Leibbrandt, Finn and Woolard (2012). To collect these measures at the level of subgroups, and combine them into the larger group, we derive the following from the definition of each inequality decomposition.

Let one category (e.g. CBSA) be indexed by j, and the subgroup category (e.g. occupation) be denoted by i. To derive the approximate overall wage inequality in each CBSA j, we calculate the following from estimated inequality measures in the subcategory (indexed by ij). Let Y denote annual wage, and y denote the logarithm of annual wage. The Generalized Entropy measures of the wage distribution, and the total variance in the logarithm of wages can be expressed as

$$\begin{aligned} GE(0)_j &= \sum_i \frac{N_{ij}}{N_j} GE(0)_{ij} + \sum_i \frac{N_{ij}}{N_j} \ln \frac{\bar{Y}_j}{\bar{Y}_{ij}} \\ GE(1)_j &= \sum_i \frac{N_{ij}\bar{Y}_{ij}}{N_j\bar{Y}_j} GE(1)_{ij} + \sum_i \frac{N_{ij}\bar{Y}_{ij}}{N_j\bar{Y}_j} \ln \frac{\bar{Y}_{ij}}{\bar{Y}_j} \\ Var(y_j) &= \sum_i \frac{N_{ij}}{N_j} Var(y_{ij}) + \sum_i \frac{N_{ij}}{N_j} \left(\bar{y}_{ij} - \bar{y}_j\right)^2 \end{aligned}$$

Section 2 reports the between-region component of wage inequality using Theil's T (GE1) as the measure for inequality. In this appendix, we show that other inequality measure yield broadly similar patterns of rising between-region inequality over time. Figures A.2 shows inequality decomposition patterns using GE(0) (Theil's L) and total variance in log wages.

A.4 Patterns in Estimated CBSA Indexes for Wage Level and Wage Spread

In Section 4, we estimated indexes for the wage level and spreads in each CBSA, based on Equations 1 and 2. In this appendix, we list these indexes across a number of years ranging the last two decades, to support the descriptive analyses in the main part of the paper. Table A.1 shows the CBSA indexes for wage level, and Table A.2 shows the CBSA indexes for wage spread (top decile to median), for the top six locations with highest indexes in each 3-year time period.

From Table A.1, top cities have always had the highest-paying IT jobs, throughout the past two decades. Silicon Valley and San Francisco, in particular, have sustained top wages in all years, where IT workers earned at least 7–14% more than in the 6th highest-paying CBSA. The very top locations also experienced faster wage growth. In recent years, a few metropolises and urban areas, including Washington DC and New York City, have become particularly attractive to local IT labor, paying the largest IT-wage premiums relative to other U.S. metroplitan areas.

On the other hand, there is almost no overlap between the lists of CBSAs in Table A.1 and in Table A.2. The places that pay least equal wages to local IT workers are not the top CBSAs that attract the most well-paid IT talent. This is consistent with Section 4.2 which finds that wage level and wage spreads are negatively correlated among U.S. cities.

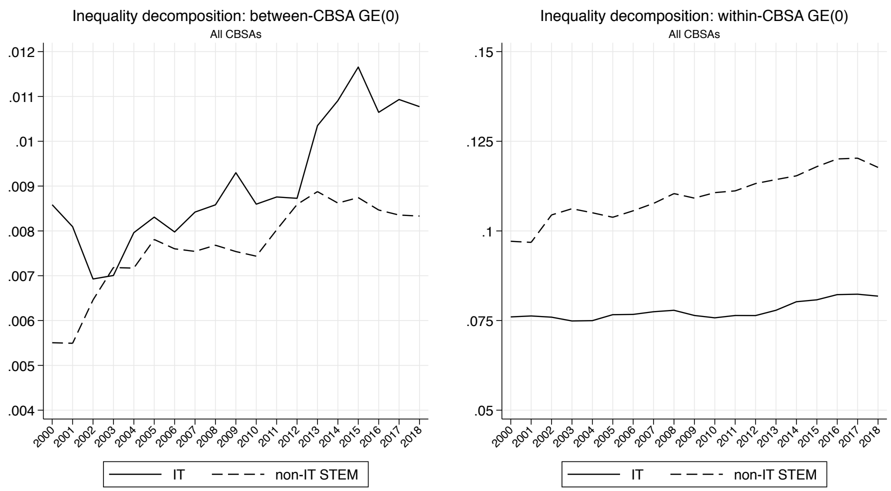
We also show the estimated occupation indexes in Table A.3 and A.4 for wage level and wage spread (top decile to median) respectively. The numbers suggested that the wage gap across locations widened substantially from 2000 to 2018. The wage difference between research scientists (highest-paying) and support specialists (lowest-paying) increased from 55% in 2000 – 2002 to 72% in 2016 – 2018. This change can potentially be explained by the skill level, as theories of skill-biased technical change explains wages growing apart between jobs that require different degrees of education. For example, computer research (and information) scientist positions typically require at least a master's degree, and they also pay the highest wages among IT occupations. Computer support specialists do not require a college degree, and they also pay the lowest wages on average.

Figure A.1: IT and Other STEM Wage Inequality (Theil's T) Within CBSAs

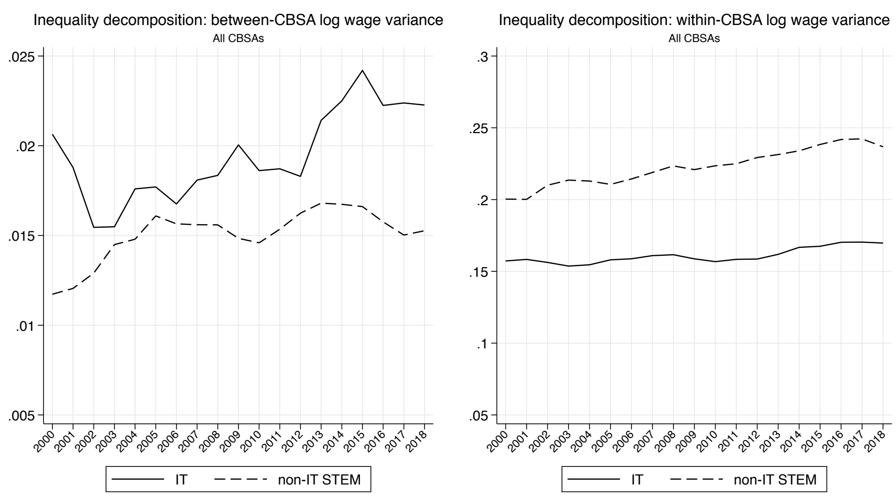
Notes: This figure shows the within-CBSA wage inequality among IT occupations (solid) and non-IT STEM occupations (dashed) respectively, in each year between 2000 and 2018. The X-axes represent each year from 2000 to 2018. The Y-axes represent the between-CBSA component of an inequality index. The inequality index is measured using Theil's T – General Entropy Index GE(1), which is more sensitive to the top part of the wage distribution, and constructed from CBSA-occupation level wage statistics (mean, 10th, 25th, 50th, 75th, and 90th percentiles). The left panel plots annual regional wage inequality for all CBSAs, and the right panel plots the same measure but excluding five areas – Silicon Valley, San Francisco, Seattle, New York City, and Washington DC.

Figure A.2: Between- and Within-CBSA Components of Wage Inequality

(a) Theil's L



(b) Variance in Log Wages



Notes: This figure presents the decomposition of wage inequality into a between-CBSA component (left) and the sum of within-CBSA components (right). It shows the decomposition among IT occupations (solid) and non-IT STEM occupations (dashed) respectively, in each year between 2000 and 2018. The X-axes represent each year from 2000 to 2018. The Y-axes represent the between-CBSA component of an inequality index. The inequality index is measured with Theil's L – General Entropy Index GE(0) in the top panel, and the variance in log wages in the bottom panel. These inequality measures are derived using aggregate wage statistics, and detailed data construction steps are outlined in Appendix Section A.3.

Table A.1: Selected List of CBSA IT-Wage Level Indexes

Rank	2000 - 2002		2006 - 2008		2011 - 2013		2016 - 2018	
1	San Jose-Sunnyvale-Santa Clara, CA	0.066 (0.025)	San Jose-Sunnyvale-Santa Clara, CA	0.079 (0.022)	San Jose-Sunnyvale-Santa Clara, CA	0.080 (0.018)	San Jose-Sunnyvale-Santa Clara, CA	0.056 (0.031)
2	San Francisco-Oakland-Hayward, CA	0.000 (/)	San Francisco-Oakland-Hayward, CA	0.000 (/)	San Francisco-Oakland-Hayward, CA	0.000 (/)	San Francisco-Oakland-Hayward, CA	0.000 (/)
3	New York-Newark-Jersey City, NY-NJ-PA	-0.071 (0.039)	Boston-Cambridge-Nashua, MA-NH	-0.049 (0.018)	Washington-Arlington-Alexandria, DC-VA-MD-WV	-0.021 (0.027)	Seattle-Tacoma-Bellevue, WA	-0.047 (0.038)
4	Seattle-Tacoma-Bellevue, WA	-0.072 (0.037)	Washington-Arlington-Alexandria, DC-VA-MD-WV	-0.058 (0.024)	Boston-Cambridge-Nashua, MA-NH	-0.058 (0.016)	Washington-Arlington-Alexandria, DC-VA-MD-WV	-0.073 (0.017)
5	Boston-Cambridge-Nashua, MA-NH	-0.074 (0.022)	Bridgeport-Stamford-Norwalk, CT	-0.061 (0.031)	Seattle-Tacoma-Bellevue, WA	-0.059 (0.013)	New York-Newark-Jersey City, NY-NJ-PA	-0.092 (0.021)
6	Worcester, MA-CT	-0.074 (0.022)	New York-Newark-Jersey City, NY-NJ-PA	-0.065 (0.014)	Durham-Chapel Hill, NC	-0.087 (0.046)	Bridgeport-Stamford-Norwalk, CT	-0.126 (0.027)

Notes: This table shows the CBSA indexes for the top six CBSAs ranked from highest to lowest – estimated from the fixed-effects regression for the level of IT wages. The name of each CBSA is listed, along with the fixed-effects estimates and standard errors. The indexes are estimated relative to one omitted location – San Francisco, and the values can be interpreted as the difference in overall IT wage levels between each CBSA and San Francisco, after accounting for the composition of occupations in each CBSA.

Table A.2: Selected List of CBSA IT-Wage Spread Indexes

Rank	2000 - 2002		2006 - 2008		2011 - 2013		2016 - 2018	
1	Fayetteville-Springdale-Rogers, AR-MO	0.217 (0.069)	Port St. Lucie, FL	0.128 (0.077)	Dallas-Fort Worth-Arlington, TX	0.143 (0.076)	Asheville, NC	0.198 (0.092)
2	Santa Rosa, CA	0.118 (0.058)	Provo-Orem, UT	0.127 (0.047)	Brownsville-Harlingen, TX	0.140 (0.086)	Provo-Orem, UT	0.117 (0.054)
3	North Port-Sarasota-Bradenton, FL	0.118 (0.061)	Santa Maria-Santa Barbara, CA	0.070 (0.030)	Provo-Orem, UT	0.134 (0.065)	Greenville-Anderson-Mauldin, SC	0.098 (0.037)
4	Houston-The Woodlands-Sugar Land, TX	0.094 (0.039)	North Port-Sarasota-Bradenton, FL	0.067 (0.058)	Springfield, MO	0.130 (0.038)	Port St. Lucie, FL	0.098 (0.044)
5	El Paso, TX	0.0932 (0.048)	Orlando-Kissimmee-Sanford, FL	0.066 (0.044)	Beaumont-Port Arthur, TX	0.119 (0.053)	Santa Maria-Santa Barbara, CA	0.098 (0.055)
6	Port St. Lucie, FL	0.089 (0.079)	Brownsville-Harlingen, TX	0.065 (0.036)	McAllen-Edinburg-Mission, TX	0.108 (0.043)	Miami-Fort Lauderdale-West Palm Beach, FL	0.084 (0.022)

Notes: This table shows the CBSA indexes for the top six CBSAs ranked from highest to lowest – estimated from the fixed-effects regression for the spread of IT wages. The name of each CBSA is listed, along with the fixed-effects estimates and standard errors. The indexes are estimated relative to one omitted location – San Francisco, and the values can be interpreted as the difference in overall IT wage spreads (top decile to median) between each CBSA and San Francisco, after accounting for the composition of occupations in each CBSA.

Table A.3: List of IT Occupation Wage Level Indexes

Rank	2000 - 2002		2006 - 2008		2011 - 2013		2016 - 2018	
1	Computer Scientists, Research	0.172 (0.017)	Computer Scientists, Research	0.304 (0.016)	Computer Scientists, Research	0.241 (0.018)	Computer Scientists, Research	0.256 (0.018)
2	Software Engineers, Systems	0.152 (0.010)	Software Engineers, Systems	0.200 (0.008)	Software Engineers, Systems	0.229 (0.007)	Software Engineers, Systems	0.195 (0.009)
3	Software Engineers, Applications	0.142 (0.008)	Software Engineers, Applications	0.157 (0.007)	Software Engineers, Applications	0.155 (0.006)	Software Engineers, Applications	0.158 (0.007)
4	Computer Programmers	0 (/)	Computer Network Occupations	0.006 (0.006)	Computer Network Occupations	0.018 (0.011)	Database Administrators	0.038 (0.008)
5	Computer Network Occupations	-0.01 0.006	Computer Programmers	0 (/)	Database Administrators	0.014 (0.007)	Computer Programmers	0 (/)
6	Database Administrators	-0.052 (0.009)	Database Administrators	-0.018 (0.013)	Computer Programmers	0 (/)	Computer Network Occupations	-0.010 (0.006)
7	Computer Support Specialists	-0.374 (0.010)	Computer Support Specialists	-0.429 (0.008)	Computer Support Specialists	-0.409 (0.006)	Computer Support Specialists	-0.459 (0.007)

Notes: This table shows the occupation indexes for all IT occupations ranked from highest to lowest – estimated from the fixed-effects regression for the level of IT wages. The name of each occupation is listed, along with the fixed-effects estimates and standard errors. The indexes are estimated relative to one omitted occupation – computer programmers, and the values can be interpreted as the difference in overall IT wage levels between each occupation and computer programmers, after controling for location fixed effects.

Table A.4: List of IT Occupation Wage Spread Indexes

Rank	2000 - 2002		2006 - 2008		2011 - 2013		2016 - 2018	
1	Computer Support Specialists	0.038 (0.009)	Computer Support Specialists	0.007 (0.006)	Computer Support Specialists	0.031 (0.007)	Computer Support Specialists	0.049 (0.008)
2	Database Administrators	0.038 (0.009)	Computer Programmers	0 (/)	Computer Programmers	0 (/)	Computer Network Occupations	0.009 (0.007)
3	Computer Programmers	0 (/)	Database Administrators	-0.007 (0.008)	Database Administrators	-0.021 (0.007)	Database Administrators	0.006 (0.009)
4	Computer Scientists, Research	-0.021 (0.019)	Computer Network Occupations	-0.037 (0.011)	Computer Network Occupations	-0.024 (0.006)	Computer Programmers	-0.003 (/)
5	Computer Network Occupations	-0.05 (0.005)	Software Engineers, Applications	-0.046 (0.007)	Computer Scientists, Research	-0.039 (0.022)	Software Engineers, Applications	-0.003 (0.007)
6	Software Engineers, Applications	-0.057 (0.008)	Computer Scientists, Research	-0.064 (0.013)	Software Engineers, Applications	-0.041 (0.007)	Software Engineers, Systems	-0.025 (0.008)
7	Software Engineers, Systems	-0.063 (0.008)	Software Engineers, Systems	-0.064 (0.007)	Software Engineers, Systems	-0.046 (0.012)	Computer Scientists, Research	-0.043 (0.016)

Notes: This table shows the occupation indexes for all IT occupations ranked from highest to lowest – estimated from the fixed-effects regression for the spread of IT wages. The name of each occupation is listed, along with the fixed-effects estimates and standard errors. The indexes are estimated relative to one omitted occupation – computer programmers, and the values can be interpreted as the difference in overall IT wage spreads (top decile to median) between each occupation and computer programmers, after controling for location fixed effects.