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
Harvard Business School

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 ‡

January 2021

Abstract

Several decades of expansion in digital communications, web commerce, and online distribution have altered regional IT labor market returns in the United States. IT occupations experienced similar wage growth as STEM occupations involving IT-related work activities, and wage inequality rose across locations and within MSAs. Supply side characteristics especially agglomeration and skilled labor pooling contributed to regional variation in IT wages since 2005. The size distribution of establishments in IT-using services industries increasingly drove IT wage inequality after 2012. While market concentration contributed to wage premiums across locations, establishment count contributed to within-MSA wage spread.

*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.

†Harvard Business School. Email: rcao@hbs.edu

‡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, and how that premium creates inequality in wages. Theories of skill-biased technical change (SBTC) offers an explanation for the premium. High 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, Levy and Murnane, 2003; Autor, 2019; Goldin, Katz et al., 2020). Viewed through this lens, the uneven deployment of the Internet in the 1990s placed additional upward 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).

If this explanation continued to have force, then wages for high skills should have been pushed higher after the turn of the millennium, when U.S. businesses made substantial investments in networking infrastructure and enterprise IT. This class of assets grew faster than any other in the nationwide economy (Byrne and Corrado, 2020). 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, placing upward pressure on a range of skilled IT labor, such as 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).

In this paper we look for evidence of a change in the premium in high skills IT wages, and analyze its causes. We characterize how similar IT occupations vary across regions over time, and analyze whether this variance is consistent with theories of skill-biased technical change (SBTC) and other explanations that stress local labor markets for skilled IT. Differences between wages for skilled occupations are persistent across locations, so our basic approach is to compare IT premiums with comparably skilled non-IT occupations, and closely examine the experience in technology hubs compared with other major US regions.

We also investigate the increasing differences between regions with the largest IT premium and others, and then analyze the factors shaping those differences. The focus on regional variance in occupational wages enables us to entertain a range of hypotheses about the factors that put upward pressures on wages.

While the role of technology hubs has received attention as a source of innovation from many authors, the comparative experience of IT workers in technology hubs, and, more broadly, regional differences between occupations, have not received more attention as an explanation for wage inequality. The differences appear to be substantial even in descriptive statistics. For example, computer and information research scientists on average pay $167,990 in Silicon Valley and $140,660 in San Francisco, at a 20–40% premium relative to Indianapolis where the same job pays around $117,260 in 2018. On the other hand, the average wages of biochemists were $115,070 in San Francisco and $105,850 in Silicon Valley, around the same amount as $118,790 which is the average wage in Indianapolis in 2018. Although this is an example especially of the tech hub's IT wage premium, the regional differences in IT wage distribution do not only pertain to Silicon Valley or San Francisco, but may be present in large MSAs overall. These facts call for explanations that involve local labor markets, and not only reasons based on the differences between-occupations.

This paper explains IT labor market inequality since the 2000s, focusing specifically on the role played by regional and occupational differences, and the underlying causes of those differences, and with most of the focus on explaining the higher end of the distribution. We collect annual wage statistics in the Occupational Employment Statistics (OES) for the 142 largest U.S. metropolitan statistical areas (MSAs) from 2000 to 2019, for each of the occupations (defined by 6-digit SOC codes) requiring 20% or more IT-related work activities. Rapid advances in information technology took place across many sectors during this time period, which also covered two economy-wide recessions – the early 2000s' dot-com bubble and the 2008 Financial Crisis.

We first establish a few stylized facts. The past two decades have witnessed slower but more broad-based wage growth for IT-skilled labor, especially compared with experience during the dot-com boom (i.e., prior to 2000). Three pieces of evidence support this view. First, wages in strictly defined IT occupational categories grew at around same rate as those in similarly skilled non-IT STEM occupations. Second, IT-specific wage premiums in tech hubs (Silicon Valley, San Francisco, Seattle, New York City and Washington DC) relative to other large MSAs did not substantially grew over time, but were sustained at around the same levels since the early 2000s. Third, wage spread and average wage at the MSA level have not been positively related after 2005, in contrast to the early 2000s where places paying higher wages to IT labor also displayed more unequal wage distribution. In other words, the top- earning IT professional were not disproportionately rewarded in frontier cities or tech hubs. These aggregate facts were somewhat surprising and not previously recognized in the literature.

We then show that IT labor regional inequality has risen over the last two decades, both in terms of wage gaps between MSAs, and within-MSA wage spreads, even after accounting for occupational wage differences and varying employment sizes across locations. Our main empirical findings focus on regional explanatory factors that determine the relative IT wage levels and spreads and changes to them. We test three distinct region-based theories around the supply and demand of talent in local IT labor markets.

We find that the explanatory power of different MSA-level factors have systematically shifted over time, again, not a pattern previously recognized by the literature. First, we find that urban agglomeration, and other local supply-side characteristics and economic conditions together explain 80% the variation in mean wages and 13% of wage spread at the MSA-level since 2005, but not in the early 2000s. In recent years, increasing total MSA-level population by 10% is associated with 1% higher IT wages, and increasing the local share of college-educated adults by 5 percentage points is associated with 1.5–1.6% higher IT wages on average. The college-educated population may be a more relevant pool for IT-skilled jobs than the overall population, as the typical education level required of most IT occupations is at least a bachelor's degree. Larger pools of IT-skilled labor match high-quality talent to firms more efficiently, and tech hubs are sustained as companies are willing to pay a premium to locate where more skilled talent is available. This confirms the theories around agglomeration economies, where supply-side factors drive up local returns to IT labor, and more abundant supply of high-skilled labor further fuel the wage advantage of some MSAs over others.

Second, we find substantial demand-side changes in the contribution of Marshallian externalities and economies of scale, especially after 2012. The primary driver of this demand-side change comes from establishments in *IT-using services* industries (e.g., finance, publishing, and business services), and not those in IT-producing or IT-using manufacturing industries (e.g., computers, electronics, and telecommunications). Also, a larger number of establishments in IT-using services industries contribute to larger wage spreads (but has no effect on average wages like in the early 2000s), especially after 2012 and this effect has increased steadily over the seven years, 2012 to 2019. These results suggest the positive association between IT-intensive establishment count and regional IT wage premium – predicted by a standard theory of Marshallian externalities – has ceased to matter in recent years, and, instead, demand for high-productivity IT labor appears to be dominated by large establishments and market concentration. This exacerbates inequality by paying higher wages in places with few overall but very large establishments, while paying lower wages where small establishments are more prevalent.

Third, we find little effect of local business dynamism, entrepreneurship and innovation on returns to IT labor. Despite theoretical mechanisms associated with spatial spillovers among inventors and concentration in IT innovation (Moretti, 2019; Forman and Goldfarb, 2020), the typical IT worker (who may not directly conduct inventive work) does not appear to earn higher wages as a result of heightened entrepreneurship and startup activity in the area. The empirical effects appear to be tiny, throughout the entire time period we examine. This result is surprising, and also is not recognized in the literature.

In summary, we shed light on several local mechanisms that drive regional variation in the wage distribution of IT workers. Our findings emphasize the view that regional experiences within local labor markets shape the returns to IT skills, and play an increasingly significant role in driving wage inequality, during the past two decades and especially since the onset of the 2008 Financial Crisis. We provide empirical evidence for how the explanatory power of different MSA-level factors have changed substantially over time, by examining our data separately in three time periods: 2000–2004, 2005–2012, and 2013–2019. Our results stem from a broad set of large MSAs across the U.S., and are not limited to tech hubs or a few locations that pay large premiums to IT-skilled workers.

1.1 Related Work

This paper contributes to several strands of literature. First, it speaks to a volume of work on the unequal effects of IT technological advances on income across locations (Forman, Goldfarb and Greenstein, 2002, 2005, 2012). In this literature, the highest income locations are also the ones experiencing the largest benefits from IT, exacerbating regional income divergence within the U.S. prior to 2000. We find somewhat distinct patterns after the end of the dot-com boom, using annual data from 2000 to 2019. The growth in the returns to labor complementary to IT capital has been rather broad-based, instead of being exceptional for a particular few occupations or in only a small number of locations. The IT wage premium in tech hubs remained at about the same magnitude of 6–12% throughout the past two decades.

Second, our findings add to the existing literature on local determinants of IT labor market returns, especially labor pooling (Cullen and Farronato, 2020), IT capital complementary (Bartel, Ichniowski and Shaw, 2007; Tambe and Hitt, 2012) and Marshallian externalities (Tambe and Hitt, 2014) that drive local wage variation through agglomeration (e.g., large supplies of skill) and distributional features of IT-intensive establishments. Our empirical results suggest that agglomeration has increasingly contributed to higher IT wages at the MSA level since 2005. We also find mixed results regarding demand-side factors, particularly IT intensity of firm establishments – while Marshallian externalities appear to drive higher wages in the early 2000s, their effects are reversed in later years. Concentration of very large establishments especially in IT-using services industries appear to contribute to higher wage levels in recent years. The number of establishments in IT-using services industries have become increasingly associated with the larger wage gaps between the higher and lower ends of within-MSA IT wage distribution after 2012.

Last but not least, our paper is related to a diverse body of work exploring the relationship between new technologies and rising concentration and inequality in the U.S. economy. This literature, however, does not emphasize *regional* variation but rather focuses on various other dimensions, e.g., decline of business dynamism (Haltiwanger, Jarmin and Miranda, 2013), IT capital and firm productivity (Tambe et al., 2020), and wage gaps between industries (Eckert, Ganapati and Walsh, 2019; Kaltenberg, 2020) and between occupations due to SBTC (Goldin, Katz et al., 2020). We stress the increasing relevance of location-specific factors in shaping the returns to IT-skilled labor, across large U.S. urban areas which experience the impact of both IT innovation and market concentration in apparent ways especially in recent years.

2 Background and Data

We start by describing the research setting and data, for analyzing the wage distribution of regional IT-skilled labor markets across the United States. We collect occupational wage statistics for each of the 142 largest metropolitan statistical areas (MSAs) and each occupation defined by 6-digit SOC codes, measured at annual frequency from 2000 to 2019. The data allows us to compare occupations in the strictly defined IT category and those in non-IT STEM occupations involving similar job tasks and activities. We match the wage data to MSA-level demographics and industrial composition, available from public sources, i.e., the Census Bureau.

2.1 Annual Occupational Wage Statistics by MSA

The main data source for this paper is the Occupational Employment Statistics (OES), published annually by the Bureau of Labor Statistics. The data contains wage statistics such as the mean, and the 10th, 25th, 50th, 75th, and 90th percentiles of the wage distribution for every 6-digit SOC occupation code in every metropolitan statistical area (MSA) in the United States.[1] The annual wages are denominated in contemporary US dollars. We collect the wage data in each year from 2000 to 2019, and harmonize the MSA codes across years to ensure we track the same geographical area over time, despite minor adjustments occasionally made to official definitions of the exact areas that correspond to some MSA codes.

The Standard Occupation Classification (SOC) is a unified system that classifies all occupations in the United States into about 900 categories, each represented by a 6-digit

[1]The OES public data sometimes contain missing values for smaller areas without sufficiently large employment sizes in a particular occupation. The data also omit quantile statistics for the top parts of the occupational wage distribution, when the wage number in the MSA for the particular occupation exceeds an upper threshold. In the latter case, we code impute the missing quantiles by multiplying the next (lower) available quantile by the average (across all occupations and MSAs in the data) ratio between the two quantiles.

SOC code. The SOC system has been twice renewed, in 2010 and 2018 respectively, updating a subset of occupation categories and assigning new codes to them each time. We use crosswalk files to match these codes across years, only for the 8 – 12 strictly defined IT occupations. For other occupations outside the IT categories, we do not attempt to map them across years more generally. The OES data does not suitably entail a panel structure, and our analyses use these wage statistically on a year-by-year basis. Even when we need to account for occupation fixed effects, we make sure to interact with the year dummies, and therefore we effectively estimate occupation-year fixed effects to stringently control for occupation-specific differences separately for each year.

The geographies studied in this paper are primarily large urban areas in the United States, where local labor markets for IT skilled professionals exist and are sufficiently large. We focus on the 142 largest metropolitan statistical areas (MSAs) in terms of total population size in 2010, and do not collect data on smaller cities and rural areas (e.g., micropolitan statistical areas). Across all these areas, the wage statistic for every IT occupation is available at least for the mean wage level in every year from 2000 to 2019. In smaller areas, wage data may be missing either because employment size was tiny or because the region was not sampled at all. A number of low-density small areas also did not have access to frontier IT infrastructure during this time (Greenstein, 2020). Since our analyses depend on having consistent data across locations on an annual basis for two decades, we only include the largest urban areas where wage statistics can be reliably measured.

To provide the reader with a more intuitive view into the wage data, Table 1 lists a few examples of wage statistics in three locations, for three IT occupations and three non-IT STEM occupations that perform IT-related tasks, at different times over the last two decades. The listed locations typically pay the highest (San Francisco CA), middle (Indianapolis IN) and lowest (Little Rock AR) IT-skilled labor wages among the 142 largest MSAs. For each location and occupation, we show the 75th percentile, the mean, and the 10th percentile of the wage distribution, as well as the relative wage ranking of the MSA, every nine years from 2000 to 2019. The top three occupations are strictly in the IT category – computer research scientists (high wage), database administrators (middle wage), and user support specialists (low wage); the bottom three occupations are in the STEM category requiring similar skills to

IT occupations – engineering managers (high wage), statisticians (middle wage), and survey and mapping technicians (low wage).

Table 1 showcases a few salient features of the data set. First, the returns to occupations differ widely across occupations. Computer and information research scientists in San Francisco pay 90% higher wages than user support specialists in 2018, for example, and this is not unique to one location or specific to IT jobs. In 2018, engineering managers in Indianapolis pay 72% higher wages than statisticians, and 181% higher wages than survey and mapping technicians. Second, the average wages differ substantially between locations. Focusing on database administrators, the mean wage level in San Francisco is 36% higher than that in Little Rock. Third, the relative wage rankings of top MSAs (e.g., San Francisco) appear persistent over time, whereas the relative wage rankings of middle and bottom MSAs can change quite a bit across years. In Section 4.3, we measure the the correlations in wage levels and spreads over time, and present these results more systematically than the illustrative examples here.

This data allows us to calculate wage inequality measures, such as Theil's indexes and total variance, for every occupation and every MSA. These standard inequality measures have the nice property of allowing analytical decomposition into between- and within-group components, for any discrete number of groups[2]. Appendix A.2 provides details into constructing wage inequality indexes based on our data, which requires approximating the wage distribution under a parametric assumption and estimating relevant parameters to fit the given data reasonably well. We present patterns in wage inequality in Section 3.

We also collect MSA-level data on demographic profiles of the population and industrial composition, as well as other miscellaneous regional variables that measure local economic conditions, innovation and entrepreneurial activities. All the data sources for MSA-level factors are listed, and construction of variables described in Appendix A.1. We discuss empirical analyses on the association between MSA-level explanatory factors and differences in IT wage distribution across locations in Sections 4 and 5.

To summarize, we outline the main data source – Occupational Employment Statistics

[2]Not all inequality indexes have this property, e.g., Gini's Index cannot be cleanly separated into additive components by group.

(OES) wage data at the granularity level of each MSA within the U.S. and 6-digit SOC code, released annually by the Bureau of Labor Statistics. We illustrate specific examples of the final data set, to motivate studying more broadly applicable features of this data.

3 Stylized Facts

This section provides stylized facts on labor market inequality in occupations requiring IT skills, across space and over the last two decades from 2000 to 2019. We examine whether the returns to IT jobs have dramatically diverged from those to other professions, in particular a subset of STEM jobs requiring similar skill sets. We also examine whether wage trends have grown apart between some locations, especially tech clusters, and the rest of the United States.

To identify the set of jobs that are comparable to those performed by IT professionals, we start with the set of Computer and Information Technology Occupations defined by the Bureau of Labor Statistics[3], which we refer to as "IT occupations". We then identify a subset of STEM occupations that require similar skills as IT occupations. To do so, we use information in O*NET data about the indirect work activities (IWA) associated with each 6-digit SOC occupation code. We include STEM occupations that share 20% or more of the same IWAs as an IT occupation.

By the criterion above, for each IT occupation there are around 12 STEM occupations on average with a substantial share of IT job tasks. As occupational codes change by year, Figure A.1 plots the annual number of SOC codes among IT occupations and similar STEM occupations separately, in each year from 2000 to 2019. In this paper, our analyses focus on annual wage statistics on these occupations only, and across the largest 142 metropolitan statistical areas (MSAs).

These stylized facts emphasize inequality in the wage distribution of IT-skilled occupations over time. Specifically, we explore whether wage levels have grown apart or closer *between* different regions within the U.S., and extent to which *MSA-specific* forces drive wage inequality patterns.

[3] https://www.bls.gov/ooh/computer-and-information-technology/.

3.1 Unexceptional IT Wage Growth Relative to STEM

We show the average trend in wage growth across IT occupations, and compare it with the part of the broader STEM labor market that requires similar skills. While IT occupations account for a large and rising share of 40 - 50% of all STEM employment, and IT technology positively affect productivity (Bresnahan, Brynjolfsson and Hitt, 2002; Bartel, Ichniowski and Shaw, 2007; Tambe and Hitt, 2012, 2014), benefits of IT do not appear to accrue to workers as drastically as they have contributed to raising productivity and market capitalization of IT-intensive firms.[4] Overall, we find little empirical support for average wage growth or wage inequality being "exceptional" among IT occupations.

Figure 1 left panel shows that both IT and similar-skilled STEM occupations have experienced steady increases over the last two decades, and that their wages grew at similar rates. In fact, wages in IT occupations increased 10 – 15% less than those in similar-skilled STEM occupations from 2000 to 2019. This runs counter to the notion that the typical IT job holder is paid much higher returns relative to other jobs, conditional on skill.

The only possibly "exceptional" part of IT wage growth, which might set it apart from similarly skilled STEM occupations, is the IT wage premium in tech hubs – a few large MSAs that pay persistently higher wages to IT workers relative to STEM jobs that require comparable skills. Figure 1 right panel shows that as early as in the 2000s, IT wages have experienced about 20% higher growth since 2000 in tech hubs – Silicon Valley, San Francisco, New York City, Seattle, and Washington DC), compared to the other 142 largest MSAs.

These patterns are consistent with anecdotes that the kind of frontier skills, often in short supply and rewarded by exceptionally high wages (Tambe, 2014), are rare elsewhere but primarily found in tech hubs, and only account for a small concentrated portion of the overall IT labor force.

Figure 2 provides evidence that rising overall wage inequality in IT-skilled occupations is not driven only by a particular few SOC codes in information technology, but rather is associated with many different occupational categories, both in IT and in STEM jobs

[4] Revenue growth were in the range of 25 – 50% for companies like Amazon, Facebook and Google on an annual basis in the 2010s.

requiring similar skills to IT. It shows the decomposition of occupation averages of the GE(1) wage inequality index into the difference between hub and non-hub MSAs (left panel) and the remaining inequality (right panel).[5] The figure made it clear that the rise in within-occupation wage inequality was particularly apparent after the 2008 Financial Crisis. It is attributable not only to the widening gap between tech hubs and the rest of the U.S., but also to wage spread after accounting for tech hub specific wage premiums.

A few hypotheses are consistent with the fact that IT wage movements track that of similarly skilled STEM wages, rather than showing distinct patterns much different from the latter. One potential contributor is that gaining functional knowledge of a new IT skill may be easy for workers who possess older (and possibly obsolete) skills. IT professionals might be substitute with workers in non-IT professions performing job tasks or work activities similar to those required by an IT occupation, e.g., dealing with electronics or conducting data analyses. Entry into IT as a profession is relatively easier than entry into occupations requiring certification and years of structured training, such as medical doctors, lawyers or tenure-track academics. Training programs and boot-camps (provided by companies such as General Assembly and Kahn Academy) indicate demand for such transition, and equip participants with necessary IT skills.

To summarize, IT-skilled labor wages have become increasingly unequally distributed over the past two decades. While regional differences (specifically between tech hubs and other MSAs) contribute partly to inequality trends, increasing within-region wage spread suggest that returns to IT skills *within* the same local labor market contribute substantially to overall inequality as well. These IT wage inequality stylized facts speak to a focused segment of recent macroeconomic trends, including rising market concentration and declining business dynamism across the U.S. (Decker et al., 2014; Akicigit and Ates, 2019).

[5]Appendix A.2 explains the details of wage inequality decomposition, which involves two steps: (1) inferring parameters associated with wage distributions within each occupation and MSA based on occupation-MSA level wage summary statistics (2) deriving aggregate inequality measures, e.g., GE(1) Theil's T and GE(0) Theil's L, by calculating weighted sums of these inferred parameters, taking into account differences in employment sizes across MSAs and occupations. The inequality measure underlying both figures 2 and 3 is the Theil's T, also known as the Generalized Entropy Index GE(1). Appendix Section A.2 describes the construction of this index, along with an alternative but closely related inequality index known as the Theil's L, or GE(0). The appealing feature of both indexes is that that they can be analytically decomposed into between-group and within-group components, where the group can be any categorical variable including occupation and region (e.g., for aggregate wage statistics), or ethnicity and gender (e.g., for individual-level salary data).

3.2 Rising IT Wage Inequality: Tech Hubs v.s. Other Large MSAs

To start understanding difference in IT labor market returns across locations, we compare wage inequality trends between workers in tech hubs and other MSAs. The growing market cap of the largest IT companies, including the Big Techs, raises the questions of whether labor wages have followed the same trends as these firms' market value. On the other hand, these companies all have headquarters and main offices in at least one of the following MSAs – San Francisco, Silicon Valley, Seattle, New York City, Washington DC. We mark these MSAs as "tech hubs" in the rest of this paper.

Figure 3 shows that wage inequality within MSAs measured by the GE(1) index have become much higher in the late 2010s relative to the mid 2000s, and that it is driven by widening wage gap between occupations (left panel) and rising within-occupation wage spread (right panel). *Between-occupation* inequality has grown steadily from 2000 to 2019 at similar rates across MSAs, consistent with the skill-biased technical change estimated to account for 38% of the increase in total wage inequality in the United States in the 21st century (Goldin, Katz et al., 2020). More surprising is the *within-occupation* wage inequality trend in IT labor markets. While IT jobs appear to be an equalizing force prior to the 2008 Financial Crisis, providing relatively equal returns to IT-skilled labor with wage differences narrowing within each occupational code categorized as information technology employment, the trend reversed itself since 2009 and the within-occupation inequality continued to rise significantly during the last decade.

We also found sharper inequality increase in some but not all tech hubs. IT wage inequality in Seattle and New York City rose much more sharply, particularly after 2009, compared to other large MSAs. This reflects both increasing wage gap between different IT occupations and larger dispersion within the same IT occupation. On the other hand, IT wage inequality trends in San Francisco, Silicon Valley and Washington DC appear similar to those in non-hub MSAs.

A few potential hypotheses might contribute to widening wage spread between workers performing similar tasks and activities within the same SOC code. Increasing market concentration may fuel firm-specific wage premiums in recent years. Job mobility and the availability of outside options may lead to persistent regional differences in IT wage inequality, e.g., non-compete enforcement vary by state.

To summarize, within-MSA IT wage inequality has been on the rise since the mid 2000s, and across large cities and not alone in tech hubs. While prominent existing theories of differences in skill level and education (SBTC) between occupations account for the phenomenon partially, we document that *within-occupation* wage spread has also increased steadily since 2009 which require alternative explanations. Regional variation in these inequality trends deserve more attention and call for further examination.

4 Wage Indexes for Metropolitan Statistical Areas (MSAs)

The purpose of this section is to derive *location-specific* wage levels irrespective of occupation, from detailed annual occupational wage statistics. This allows us to present clean MSA-level evidence each year from 2000 to 2019, such as whether the wage premium in particular locations (e.g., tech hubs) are persistent across time, or the correlation between the mean and the spread in the local IT wage distribution, after taking other factors into account especially differences in employment sizes and wage levels across occupations.[6]

The previous section shows aggregate wage inequality trends, which did not account for regional differences in employment size. For example, the relative employment ratio of computer and information research scientist to programmer may be higher in New York City than in Indianapolis, and computer research scientist positions on average pay higher wages than programmer positions. Therefore, the higher average IT wage in New York City reflects *both* a larger absolute city-specific wage premium *and* a larger employment share in the highest-paying occupation (i.e., computer and information research scientist). We need more than descriptive statistics (over average wages) to compare location-specific wage premiums between one MSA and another.

The labor markets for IT in some locations (e.g., tech hubs) are particularly more attractive than for other types of similarly skilled labor. In this section, we also estimate an IT-specific

[6]The analyses in section isolate contribution *specific to* each MSA, and derive this contribution as an MSA-level index. But at the same time we produce similar indexes for each occupation, and show results on the occupation-specific indexes in Appendix A.3.

wage premium in particular MSAs, and extend our comparison in Section 3 of trends between tech hubs and other large MSAs. We can identify this IT-specific wage premium in each year, because our data include non-IT STEM occupations that require relevant IT skills.

4.1 IT-Skilled Wage Decomposition: A Hedonic Regression Approach

We outline the empirical approach to estimating MSA wage indexes, by isolating location-specific contribution to aggregate wage statistics in each year from 2000 to 2019. This approach accounts for other factors that also drive overall wage inequality, especially the differences in the occupational composition of employment in each MSA.

We estimate hedonic regressions that decompose a wage statistic in a given year into a composite of indexes associated with all MSAs and all 6-digit SOC codes relevant to IT-skilled occupations, shifted by a constant wage base across occupations and locations. These indexes are estimated as fixed effects (of every occupation and MSA) in equation 1, one at a time for each wage statistic (including the mean, and the 10th, 25th, 50th, 75th, and 90th percentiles) and year separately. The data for these regressions consist of occupational wage statistics for typical IT-skilled workers, i.e. it includes all STEM occupations sharing at least 20% of indirect work activities (IWAs) of at least one 6-digit SOC code strictly in the IT category.[7] In addition, we exclude occupations with 90% or a higher percentage of missing wage data among all 142 MSAs. This step drops esoteric occupations for which not enough observations are available to estimate robust occupation fixed effects.

Because the observations contain non-IT occupations (but which require a subset of tasks and activities typically performed by IT professionals), we can estimate an IT-specific wage premium for *every MSA*. However, the paper focuses on estimating the IT-specific premiums for only tech hub MSAs[8], because these MSAs appear to be primary drivers of an IT-specific premium, and otherwise IT wage trends are broadly convergent with similarly skilled STEM labor, as illustrated in Section 3. In the spirit of first order approximation, we omit the interaction terms between occupation and other locations, except for each of the five tech-

[7]The IT category contains all of the occupations in the list from https://www.bls.gov/ooh/computer-and-information-technology/. All occupations in this list are also classified as part of STEM, according to the Bureau of Labor Statistics (https://www.bls.gov/oes/topics.htm#stem).

[8]We can estimate this premium for every MSA, but since such premiums are primarily sustained in tech hubs rather than elsewhere, we focus on estimating them for tech hubs in the main paper.

hub MSAs, which appear to pay persistently higher wages for IT labor than STEM labor conducting similar work tasks and activities.

Equation 1 specifies the exact hedonic regression for decomposing IT-skilled wages. Subscripts for wage statistic (e.g., 25th percentile) and for year are omitted to simplify notations.[9] In practice, w also need to specify a base category each among MSAs and occupations, and the estimated fixed effects I^{msa} and I^{occ} measure the relative wage levels of other MSAs and occupations as distances from the base category. We choose San Francisco (code 41860) as the base MSA category, and Computer and Information Research Scientists (2010 SOC code 15-1111) as the base occupation category.

$$\log(W_{ij}) = I_i^{occ} + I_j^{msa} + \mathbb{1}\left(j \in \{SV, SF, SA, NY, DC\}, i \in ITOcc\right) I_j^{hub} + \epsilon_{ij} \tag{1}$$

The outcome variable $\log(W_{ij})$ denotes the logarithm of annual wage typical to 6-digit SOC code i and MSA j. The set $ITOcc$ consists of all 6-digit SOC codes strictly defining the IT category of occupations, which are between 8 – 12 codes depending on the year. Estimates for MSA-specific fixed effects I_j^{msa} are derived as MSA wage indexes, measuring location-specific wage difference between MSA j and San Francisco (i.e., the base category). Estimates on the interaction term, i.e. the binary indicator of i being a strictly defined IT occupation and j being a tech hub, are denoted by I_j^{hub} but only estimated when j is one of the tech hub MSAs, and reflect the IT-specific wage premium in MSA j relative to non-IT STEM occupations involving similar tasks and activities to those of IT professionals.

Wage indexes are estimated not only for the mean statistic, but also for different quantiles of the distribution. This allows us to further derive MSA wage spreads, by calculating the difference between an MSA index for a higher-quantile statistic and that for a lower-quantile statistics. In the equation below, H denotes the higher quantile (e.g., 75th percentile) and L denotes the lower quantile (e.g., 10th percentile).

$$WageSpread_j^{msa}(H, L) = I_j^{msa}(H) - I_j^{msa}(L)$$

[9] Here and in the rest of the paper, shorthand notations for each of the five tech hubs are as follows: SV refers to Silicon Valley, or San Jose-Sunnyvale-Santa Clara CA (MSA code 41940); SF refers to San Francisco, or San Francisco-Oakland-Fremont CA (MSA code 41860); SA refers to Seattle, or ; NY refers to New York City, or ; and DC refers to Washington DC, or .

Here are examples to make the interpretations of the estimated indexes more intuitive. In 2013 (or pick any year), a programmer in New York City at the 75th percentile of the local wage distribution was paid a composite wage of $I^{occ}_{PGM}(0.75)+I^{msa}_{NY}(0.75)+I^{hub}_{NY}(0.75)+Const$. In the same year, a programmer in Indianapolis at the 10th percentile of the local wage distribution was paid a composite wage of $I^{occ}_{PGM}(0.1) + I^{msa}_{IA}(0.1) + Const$ where IA denotes the MSA specific to Indianapolis. The constant term $Const$ reflects the estimates of logarithm of the wage level in the base categories, i.e., typical wage of a computer and information research scientist in San Francisco.

To summarize, we have outlined a methodology to isolate *location-specific* wage components, for the set of STEM occupations (including the IT category strictly) that perform tasks and activities similar to those of IT professionals. For any given MSA and wage statistic, we can use a hedonic regression to estimate MSA-specific wage indexes, as well as IT-specific wage premiums in each tech hub, within the same regression specification. This methodology also allows us to derive location-specific *wage spread* indexes, by subtracting the wage index for a lower quantile (e.g., 10th percentile) from that for a higher quantile (e.g., 75th percentile).

4.2 IT Wage Premiums in Tech Hubs

We explore whether tech hubs pay IT-specific wage premiums relative to non-IT STEM occupations, and whether the magnitude of these premiums, if any, changed over time since the 2000s. Tech hubs are headquarters to many powerful digital companies with large and expanding market cap, or locations in which these companies operate heavily.[10] We suspect wage trends in IT categories may have been different in tech hubs, even if they are broadly similar to non-IT STEM wages requiring related skills in most MSAs, as shown in Section 3. To examine this case, we plot the estimates on the interaction terms I^{hub}_j for each of the five hub MSAs $j \in \{SV, SF, SA, NY, DC\}$ in Figure 4, for the mean wage statistic in each year from 2000 to 2019.

Unsurprisingly, we find statistically significant IT-specific wage premiums (i.e., above-zero interaction fixed effects) in all of the five tech hub MSAs. These premiums reflect the

[10] A few examples are Alphabet (Google), Amazon, Apple, Facebook, Microsoft, Netflix.

wage difference between IT workers and similarly skilled STEM workers in the same local labor market, and are rather persistent across time. Since 2010, Silicon Valley and Seattle have sustained the largest premiums of 0.09 – 0.15 among hub MSAs, which can be interpreted as IT workers in the same location earning 9 – 15% more than similarly-skilled workers in the STEM labor market. This is a sizable advantage of tech hubs over the rest of the large MSAs in the United States.

Somewhat more surprisingly, these wage premiums paid to IT workers in tech hub MSAs did not increase at all that much over time, despite the rapid expansion and revenue growth of tech companies that are based in these locations. Instead, the IT-specific wage premiums either slightly decreased or remained about the same since 2015, in almost all of the tech hubs. This is true for higher quantiles of the wage distribution, and not only for the mean statistic. The estimated coefficients $I_j^{hub}(q)$ based on the 75th percentile ($q = 0.75$) and the 10th percentile ($q = 0.1$) wage statistics show similar patterns to those in Figure 4, for each j as any of the tech hub MSAs. The lower and higher parts of the IT-skilled wage distribution within tech hubs do not appear to deviate drastically from the average wage level in the local IT labor market.

To summarize, the wage growth experienced by IT professionals has not been exceptional relative to similarly skilled STEM labor since the 2000s, even within tech hubs where highly productive firms and local employment opportunities are most abundant among all of the large cities in the United States. Contrary to a superstar theory in the *labor market* (not necessarily for the product market), the highest-paying occupations and the top parts of the wage distribution did not experience the kind of rapid growth in returns to skills that would set them apart from the rest of the IT-skilled labor market.

4.3 Metropolitan Statistical Area (MSA) Wage Indexes

We show that MSA-specific variation in IT-skilled wages are stable and not changing over time, and explore both the average wage level and wage spread within MSAs. To do so, we collect the MSA wage indexes estimated from equation 1 across years and wage statistics, and draw the transition plots for indexes in Figure 5. These indexes reflect the relative advantages of local IT-skilled labor markets in different geographic areas in the United States, in any given year from 2000 to 2019.[11]

The top panel of Figure 5 focuses on the MSA *mean wage* indexes, and shows the transition from 2000–2004 to 2005–2012 (left), and from 2005–2012 to 2013–2019 (right). The *mean* wage indexes are highly persistent over time, as the scattered data points are closely aligned with the 45-degree line. The relative magnitudes of wage premiums specific to the top wage MSAs did not drastically change over time. The SV/SF mean wage indexes have been 6–8% higher than the next highest-paying MSA throughout the last two decades.[12]

The bottom panel of Figure 5 focuses on the MSA *wage spreads*, and show the transition from 2000–2004 to 2005–2012 (left), and from 2005–2012 to 2013–2019 (right). Specifically, the wage spread for each MSA is calculated as the difference between the 75th percentile and the 10th percentile wage indexes. These wage spreads appear rather persistent over time. The scattered data points in the wage spreads transition plots follow a roughly linear relationship, and are close to but shifted to the left of the 45-degree lines.

To summarize, IT-skilled wage inequality did not drastically alter among large MSAs across the United States in the past two decades. The MSAs where returns to IT-skilled labor were the highest sustained the advantage in paying high wages across years. On the other hand, locations with more unequal IT labor markets in the 2000s continued to have large wage spread at present. Interestingly, the locations with the largest IT-skilled wage spread have primarily been outside tech hubs, including a few large MSAs in Texas such as Houston, San Antonio, and El Paso. In these locations, relatively low average wages but large dispersion in returns to IT labor potentially suggest unsaturated demand for IT talent relative to skill supply. Differences across locations are not random, but appear to persist over time, calling for region-based theories to explain the advantages sustained by some local labor markets over others, which we discuss in Section 5.

[11]Full lists of MSA wage indexes for all 142 MSAs by year and wage statistic are available upon request – please email the authors.

[12]Not only SV/SF, but a few other cities in California are among the top wage MSAs. Wage level indexes may reflect differences in living costs between different MSAs.

4.4 Relationship Between Within-MSA Wage Levels and Spreads

We explore the correlations between wage levels and wage spreads at the MSA level as well as their changes over time, to reveal additional evidence of broad-based changes in wage inequality among IT-skilled workers. Our findings suggest region-based theories for determining IT labor market returns, and point to further examining the extent to which MSA-specific factors drive wage differences across locations.

Across all of the 142 largest MSAs, the *mean wage* indexes do not correlate strongly with *wage spreads* in a given location. Appendix Figure A.2 plots the correlations between the average wage index and the 75th-to-10th percentile index spread among MSAS in 2000–2004 (left) and in 2013–2019 (right). Tech hubs paying the largest wage premiums are not the same locations as those with the most unequal IT labor markets, and the wage spreads in SF/SV have been in the middle range among large MSAs. This is additional evidence against "exceptional" returns to IT skills, and rejects the superstar theory in U.S. IT-skilled labor market which would have predicted a positive relationship between absolute levels and dispersion in returns to IT labor.

On the other hand, locations outside tech hubs have become less unequal (relative to San Francisco) over time. The bottom panel of Figure 5 shows transition plots for within-MSA 75th-to-10th-percentile wage spreads from 2000–2004 to 2005–2012 (left) and from 2005–2012 to 2013–2019 (right). In both transition plots, the majority of MSAs are situated on the right side of the 45-degree lines. The number of MSAs with IT labor markets more unequal than San Francisco has decreased, from the early 2000s to more recently, as the position of the Y-axis (relative to the mass of the data points representing MSAs) have shifted up in the right panel compared to the left panel in Appendix Figure A.2.

To summarize, our findings suggest no relationship between wage level and wage spread, across all large MSAs and over time, echoing stylized facts in Figure 2 and 3 suggesting broad-based changes in IT-skilled wage inequality trends, which are not entirely explained away by tech hubs. However, there are substantial difference across locations in IT wage inequality trends, emphasizing the need for region-based theories to further explain the roles of location-specific factors in determining IT-skilled labor market returns. These differences persist after controlling for between-occupation wage differences, hence standard theories in the literature such as SBTC do not sufficiently explain them.

5 Regional Explanations for IT-Skilled Wage Inequality

The purpose of this section is to assess the contribution of region-specific factors in shaping IT labor market returns, and driving the variation among different local labor markets at the MSA level across the United States.

Previous section showed that there are persistent differences in wage trends between local IT labor markets in different parts of the United States, and not simply attributable to a few tech hubs but broad-based inequality trends across many large MSAs. We present empirical tests for a few region-based theories that potentially shape the distribution of returns in local IT labor markets, and map these theories separately into in supply-side and demand-side variables that driving regional variation in wage levels and inequality in local IT labor markets at the MSA level.

5.1 Estimating the Effects of MSA-Level Explanatory Factors

We provide an empirical framework for testing region-based theories explaining the variation in IT labor market returns across MSAs. We estimate the coefficients on MSA-level explanatory factors in an OLS regression, with log wage levels as the outcome variable across all of IT occupations (6-digit SOC codes) and the 142 largest MSAs in the United States. The explanatory factors correspond to region-based theories related to both the supply and demand side of the IT labor market, which we describe in Section 5.2.

In Equation 2, $MSAExplanatoryFactors_{j,t-1}$ denote regional explanatory factors in MSA j, measured at time $t-1$, lagged one year from the wage statistics W_{ijt}. These factors are measured a year prior to wage statistic, so that they were not being simultaneously determined by other variables shaping the distribution of IT wages in the same time period.

We construct two outcome measures based on a subset of wage statistics, which cover both the typical MSA wage level (the mean), and the MSA wage spread (difference between the 75th percentile and the 10th percentile of the IT wage distribution). Equation 2 is estimated for the mean wage statistic.[13]

$$\log\left(W_{ijt}^{Mean}\right) = \beta \cdot MSAExplanatoryFactors_{j,t-1} + \rho \cdot Controls_{j,t-1} + \xi_{it} + \epsilon_{ijt} \tag{2}$$

We estimate the equation on pooled data across three time periods, which we refer to as "the early 2000s" (2000–2004), "around financial crisis (2005–2012)", and "post 2012" (2013–2019). The data consists of annual occupational wage statistics for 6-digit SOC codes that are in strictly-defined IT categories. The regression controls for occupation-year fixed effects ξ_{it} at the 6-digit SOC code level, as well as demographics $Controls_{j,t-1}$ at the MSA level and measured a year before.

We also study regional determinants of wage spread with equation 3, defined by difference between the 75th percentile and the 10th percentile of the distribution of the logarithm of MSA occupational wage, to shed light on the region-based theories explaining variation in wage inequality across time and locations.

$$\log\left(\frac{W_{ijt}^{P75}}{W_{ijt}^{P10}}\right) = \gamma \cdot MSAExplanatoryFactors_{j,t-1} + \kappa \cdot Controls_{j,t-1} + \eta_{it} + \nu_{ijt} \tag{3}$$

We examine changes in the contribution of each regional factor over time toward overall wage difference across locations, by estimating equations 2 and 3 for different time periods. The time trends can shed light on potential demand-side explanations for the reversal in wage inequality trend in the more recent time period after 2009, which we presented in Section 3 Figure 3.

The approach also allows us to explore whether changes in IT wage inequality trends were attributable not only to the tech hubs – a handful of leading cities with head offices and major operations of highest market cap corporations in IT-intensive industries, but also to large cities outside the tech hubs more generally. For the region-based theories to hold, we should find evidence consistent with economy-wide effect of the explanatory factors contributing to wage trends outside tech hubs.

[13]The same equation can be estimated for every wage statistic (e.g., the mean, and the 10th, 25th, 50th, 75th, and 90th percentiles), but since the goal of 2 is to understand what drives overall IT wages in a given MSA, we use the mean wage statistic. We also construct wage spread measures based on differences in quantile statistics, in a separate regression.

5.2 Supply- and Demand-Side Factors Shaping Local IT Wages

We organize region-based explanations of variation in IT wages around several theories, which primarily concern either the supply side or the demand side of the IT labor market. We discuss these theories in detail, and map them to relevant MSA-level characteristics to test their empirical relevance. The main supply-side theory concerns agglomerative spillovers, typically measured by population variables. We focus on both total population and the share of high-skilled labor (proxied by college-educated adults). Two primary demand-side theories are related to IT capital complementarity and business dynamism respectively. We focus on MSA-level variables measuring the composition of industries and firm dynamics to test for the former, and new business applications and Kaufmann indexes to test for the latter.

Agglomeration Economies and Labor Pooling. This theory 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 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 MSA-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 MSA 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 MSA.

IT Capital Complementarity and Marshallian Externalities. In growth theory, IT capital and labor are complements in the production function, and firms investing more heavily in capital raise the productivity of complementary labor and hence wages. Wages may be higher for IT employees at firms that use IT capital more intensively, because it is complementary to the workers that either use IT to perform their tasks or product IT goods and services. Hence we focus on IT-intensive firms (deploying IT capital intensively), which can be further classified into IT-using and IT-producing (and possibly both) categories.

Without establishment-level measures of IT capital stock, we rely on industry composition at the MSA level to measure aggregate dynamics of IT-intensive firms in a given location. We use County Business Patterns (CBP) data to measure the number of establishments in each MSA and industry at the 5-digit NAICS code level. County Business Patterns data are reported at an annual frequency from 1999 to 2018, and contain the establishment counts in detailed size buckets ranging from under 5 to over 5,000 in most years.[14]

We follow the definitions for IT-intensity (separately for using and producing IT) in the literature, to refine the subcategories for high IT-intensity industries. For IT-producing industries, we use classic definitions based on (Jorgenson et al., 2005) and (Forman, Goldfarb

[14]This is not true before 2003, when in each year the top size bucket had a lower threshold of 1,000.

and Greenstein, 2012), and add communications to the list. We consider all industries with medium-high or high digital intensity in Calvino et al. (2018) as IT-using industries, and further divide these IT-using industries according to whether they are in the manufacturing or services sector. A large share of platform companies that experienced soaring productivity growth in recent years are in IT-using services industries (e.g., Amazon, Uber, AirBnb among others), and they are quite distinct from many traditional firms in the manufacturing sector, and hence we separately derive the IT-intensity measure for services and manufacturing.[15]

Large firms may 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 measures for the relative dominance of large establishments in IT-using and IT-producing sectors respectively. To derive measures of MSA-level IT market concentration or dominance of large establishments, we calculate an approximate Herfindahl–Hirschman Index based on employment size distribution within MSAs as follows.

Denote the total number of employment size buckets as K, and denote the specific size buckets by their endpoints $1 = B_1 < B_2 < \cdots < B_K$, and the number of establishments in either IT-producing (p), IT-using manufacturing (m) and IT-using services (s) with employment sizes within each bucket $B_k \leq Employment < B_{k+1}$ in MSA j as n^{θ}_{jk} where θ denotes IT-intensity type and $\theta \in \{p, m, s\}$. We compute the approximate HHI measure for

[15]We classify International Standard Industrial Classification (ISIC) Rev. 4 codes classification for industries, and mapped the set of IT-producing and IT-using industries defined in ISIC into 5-digit NAICS codes using a crosswalk file from the Census Bureau, and then match with County Business Patterns data on industry composition and firm dynamics at the MSA level. The list of IT-producing industries include: Computer, electronic and optical products (26), Telecommunications (61), and IT and other information services (62-63). The list of IT-using manufacturing industries include: 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), and Furniture, other manufacturing, repairs of computers (31-33). The list of IT-using services industries include: 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 (94-96).

each MSA and IT-intensity type as

$$HHI_j^{\theta} = \frac{\sum_{k=1}^{K} n_{jk} B_k^2}{\left(\sum_{k=1}^{K} n_{jk} B_k\right)^2}$$

We use approximate employment shares of establishments within IT-intensive industries to construct the concentration index[16], separately for each of the three aggregate categories – IT-producing, IT-using manufacturing, and IT-using services, respectively. We normalize the concentration index to have variance equal to 1 every year, across the 142 MSAs in our data.

Innovation, Business Dynamism, and Entrepreneurship. This set of explanations stresses variance in regional innovation, high-growth entrepreneurship, and business dynamism. Invention tends to occur in geographically concentrated clusters, measured by aggregate venture capital and patenting activities (Kerr and Robert-Nicoud, 2020). 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).

We proxy for inventive activity within a MSA using the total number of Computers & Communication patents assigned to inventors in the MSA (Forman and Goldfarb, 2020). 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.

On the other hand, 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

[16] Exact establishment-level employment sizes are not observed, but aggregate counts within ranges of employment sizes are available. Also, data is not available for constructing alternative measures, e.g., using annual revenues or payroll.

highly productive, and experience transformative growth and have a disproportionate impact on productivity and employment in the economy.

To measure business dynamism and entrepreneurial activity, we use Kauffman Early-Stage Entreprneurship (KESE) Index (published by the Kauffman Foundation) and newly registered businesses (available as part of Business Formation Statistics tracked by Census Bureau). The data are both aggregated to the state level and available from 1999 to 2018.[17] The Kauffman (KESE) Index measures total quality-weighted entrepreneurial activities within a region. More weight is put on high-growth "transformative" entrepreneurship in the calculation than other types of new firms (which may include small businesses that do not grow in size).[18]

The patent data are available for each MSA, while the Kauffman and BFS data are available only at the state level. In only a few large states (specifically, California and Texas), the entrepreneurship indicators are coarser than ideal because there are more than a few large MSAs with varying labor market characteristics and business formation dynamics. However, most states have very few MSAs 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 MSAs for most of our data.

To summarize, we organize the empirical analyses around three sets of theories that explain location-driven variation in IT wage levels and inequality. Each theory relates to either the supply or demand side of local labor markets, and we test them by identifying measurable MSA-level variables corresponding to each theory. We assess these regional factors' influence on wage level and inequality in the local IT labor market. For more details on data sources and variable construction, see Appendix Section A.1.

[17]Public download information can be accessed at https://indicators.kauffman.org/ for the Kauffman Early-Stage Entreprneurship (KESE) Index and https://www.census.gov/econ/bfs/index.html for Business Formation Statistics (BFS).

[18]More specifically, the Kauffman (KESE) Index 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.

5.3 Empirical Results and Discussion

We present the empirical results from estimating equation 2 and 3 on three time periods: "the early 2000s" (2000–2004), "around financial crisis (2005–2012)", and "post 2012" (2013–2019), to contrast the explanatory power of different theories and explanatory factors over the last two decades at different times. Empirical evidence points to pronounced changes in the influence of supply- and demand-side regional characteristics on both the level (mean wage) and inequality (wage spread) of IT labor markets at the MSA level.

Before presenting regression results, we summarize explanatory factors on an annual basis from 2000 to 2019 in Figure 6. The figure plots the top decile, median, and bottom decile values in each of the explanatory variables (at a 1-year lag). The summary statistics of these variables remain overall stable over time, and fluctuate within reasonable ranges from year to year. Regression results are organized by outcome (i.e., mean wage and wage spread) and time period (i.e., pre-2005, 2005–2012, and post-2012) in Tables 2 – 7.

Urban agglomeration has been a leading explanation among supply-side factors driving up local returns to IT labor since 2005, and especially more abundant supply of high-skilled labor further fuel the wage advantage of some MSAs over others. On average, a 10% increase in total MSA-level population is associated with about 1% higher IT wages, and a 5 percentage point increase in the share of college-educated adults is associated with 1.5–1.6% higher IT wages on average, based on column 8 of Tables 2–3. This contrasts with the early 2000s, when total population appeared to have a negative association with average IT wage and supply of high-skilled workers only a weak and not statistically significant effect, after controlling for the demand-side counts of IT-intensive establishments (Table 4 column 8). These evidence suggest agglomeration increasingly contribute to IT labor productivity and explain higher local wage premiums, especially since the onset of the 2008 Financial Crisis.

These findings are consistent with supply-side theories especially labor pooling, where a larger local pool of IT-skilled workers facilitates less frictional matching between firms and potential hires (Cullen and Farronato, 2020). Most IT professions require at least a college degree, and hence the share of college-educated adults is another indicator of the size of the pool of IT-skilled labor supply, and appears associated with higher returns to IT jobs after accounting for total population. Tech hubs are sustained, as firms choose to locate where high-quality talent concentrates in the local area, in order to tap into the local labor market for potential hires and access productive workers more easily.

Supply-side factors and other local demographics[19] have become relatively more important in driving average wage levels since the 2000s, but their explanatory power of wage spread fell over time. MSA-level demographics explain over 80% of the variation in log wages among IT occupations since 2005 (R^2 in column 1 of Tables 2 and 3), and around 75% in the early 2000s (R^2 in column 1 of Table 4). On the other hand, the same MSA-level demographics explain around 13% of the variation in IT wage spreads post 2012 (R^2 in column 1 of Tables 5), around 15% between 2005 and 2012 (R^2 in column 1 of Tables 6), and around 20% in the early 2000s (R^2 in column 1 of Tables 7).

Switching focus to the demand side of the local IT labor market, we find particularly striking evidence for wage polarization in locations with more IT-intensive establishments and higher concentration of establishment-level employment. This holds in the regression results post-2012 (Tables 2 and 5), in contrast with a set of completely opposite results before 2005 (Tables 4 and 7). These results together suggest a reversal in the distributional role of IT-intensive firms that have determined the features of local demand for IT-skilled labor over the past two decades. While Marshallian externalities could explain the positive relationship between IT-intensive establishment count and average IT wage level in the early 2000s, this association ceased to exist in later years.

While booming IT-intensive sectors raised the overall returns to IT labor in the early 2000s and the gains appeared to have been distributed across both smaller and larger establishments, nowadays returns to IT labor in places with heavy presence of IT-intensive sectors appear to be dominated by large establishments, fueling inequality by paying higher wages where there are few overall but very large establishments (Table 2), while paying

[19]The set of control variables include regional demographic characteristics, constructed from the American Community Survey (post-2005) and U.S. Census (2000) data. These variables include The control variables include local economic conditions such as median income, unemployment rate, the share of population below the poverty line, population shares in each of the following age groups: 20 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and above, as well as the population shares in each of the following ethnic groups: Caucasian, African American, Native American, and Asian. Annual statistics on MSA-level demographics were not available between 2000 to 2005, and therefore we use the 1999 Census data to fill the missing values. For all other years, we use the annual demographics with a one-year lag.

lower wages in locations with many smaller establishments (Tables 2 and 3).

A one standard deviation increase in concentration (measured approximately as employment HHI) is associated with at least 0.8% higher IT wage, while cutting total establishment counts by half is associated with at least 3.5% higher IT wage on average after 2012, both statistically significant. In contrast, doubling the number of establishments pre-2005 was associated with 4.8% higher IT wage, while a one standard deviation increase in the concentration of large establishments did not alter the average wages but raised the top part of the wage distribution by about 1%. In addition, locations with a larger number of IT-intensive establishments have also been associated with larger wage spread post-2012 (Table 5).

We present more detailed evidence by further separating IT-intensive sectors into IT-producing, IT-using manufacturing, and IT-using services industries, and show the magnitudes and standard errors of coefficient estimates of the effects of establishment count (Figure 8) and concentration (Figure 7) in each year from 2005 to 2019. This is the time period when empirical evidence appears to run counter to a simple theory of Marshallian externalities predicting higher wages to be associated with larger clusters of IT-intensive firms.

The rising impact of concentration in recent years appears to be primarily associated with *IT-using services* industries, as suggested by panel (a) of Figures 7 and 8, and not so much with IT-producing or IT-using manufacturing industries. The trend was more pronounced after the 2008 Financial Crisis and in more recent years. Larger establishment counts in IT-using services have been associated with significantly larger wage spreads especially in recent years since 2013, and the effect sizes have steadily increased year by year, as illustrated by the last subfigure in Figure 8 panel(b).

Innovation and entrepreneurship appear to drive both wage level and inequality in local IT labor markets. The concentration of highly skilled workers in an urban location generates knowledge and productivity spillovers between closely situated workers. Inventors tend to form smaller clusters than industrial workers in general, implied by smaller radius of spillover effects that decay faster across space for innovation activities than other types of industrial activities broadly defined (Carlino and Kerr, 2015; Moretti, 2019). Also, IT patenting activities have become more geographically concentrated in recent years (Forman and Goldfarb, 2020). IT innovation may also raise labor productivity and raise wages of workers in complementary labor markets, in addition to workers directly engaged in innovation (Kline et al., 2019). Entrepreneurship may also worsen income inequality more broadly, and not specifically in the IT labor market (Marinoni, Voorheis et al., 2019). Empirically, however, the impact of factors related to entrepreneurship and IT innovation appear to be tiny. A one standard deviation increase in log number of new business applications is associated with 0.41% higher mean wage and 0.5 percentage point increase in the 75th-to-10th-percentile wage spread, but both statistically significant at the 1% level post-2012.[20] The coefficients on IT-related (Computers & Communication) patents are positive but very small, and not statistically significant.

We also note that the fact that wage level does not appear to be positively associated with wage spread since 2005 (Tables 5 and 6), echoing descriptive figures in Section 4.3 showing that MSA-specific indexes for wage level and spread appear uncorrelated. On the other hand, the positive and statistically significant coefficients on log mean wage in Table 7 suggest that MSAs paying relatively higher average wages also used to provide even larger returns to IT-skilled labor at the higher end of the wage distribution, in the early 2000s. This, however, ceased to be the case during and after the time period around the 2008 Financial Crisis.

Last but not least, we show that results are very similar even when we drop tech hubs – Silicon Valley, San Francisco, Seattle, New York City, and Washington DC, from the data. The results after excluding the tech hub MSAs are presented in Appendix Tables A.1 – A.6 and Figures A.4 – A.5. This suggests that the effects of regional factors on IT wage level and inequality identified in supply- and demand-side theories are not driven by a specific few MSAs, but rather reflect broad-based mechanisms across the economy. This is somewhat surprising, as frontier advances in IT technology post-2012 (e.g., artificial intelligence and big data) have been concentrated in a very small number of locations. Somewhat more aligned with these results is the mechanism in Baqaee and Farhi (2020), where aggregate increasing returns to scale are linked to changes in local population (e.g., due to natural and policy causes such as rising immigration or declining fertility), rather than an improvement

[20]Business Formation Statistics data are not available before 2005, and therefore we use the Kauffman (KESE) Index to measure entrepreneurship in the early 2000s instead.

in technical efficiency. Increased market size triggers improvement in resource allocation, but through shifting resources from smaller firms into the high-markup larger firms which are also more productive.

To summarize, three key empirical findings emerge from the analyses in this section. First, population size and skilled labor supply explain a large portion of IT wage differences across MSAs, and confirm theories of agglomeration economies such as labor pooling and productivity spillovers. Second, economies of scale and competition dynamics in IT-intensive (and especially IT-using services) sectors have been increasingly relevant factors driving regional inequality in returns to IT labor. Third, gains to entrepreneurship and IT innovation provide little generalized benefit to typical IT workers, or at least not in the form of regular wage compensation. These results are not driven solely by tech hubs or a few locations rewarding large premiums to IT skills, but reflect broad-based evidence for regional determinants of returns to IT labor across the United States.

6 Conclusion

The paper stresses the role of local factors that shape returns to IT-skilled labor over the past two decades. We start with descriptive patterns, analyze MSA-specific differences in wages of IT-skilled occupations, and examine the relative importance of various region-based theories and corresponding MSA-level explanatory factors in shaping local IT wage level and inequality across time.

Descriptive evidence rejects IT-exceptionalism which characterizes the growth in returns to IT skilled labor as particularly explosive for typical IT professions, and deviating from long term trends for similar non-IT occupations. Instead, IT occupational wages have followed very similar trajectory as those in similarly skilled STEM occupations. Also, IT wage trends appear to be driven by economy-wide and broad-based mechanisms, not entirely explained by extraordinary supply and demand dynamics in a few tech hubs.

Returns to IT-skilled labor have become more unequal over time, especially since the 2008 Financial Crisis. IT wage inequality increased both within the same local labor market, and between different regions across the United States, even after accounting for existing theories such as SBTC that emphasize wage gaps between occupations and skill requirements. This calls for addressing a gap in the literature, where attempts to explore factors contributing to regional variation in IT wage inequality are dearth.

We explore a few region-based theories that potentially explain the variation between IT labor markets in different locations: agglomeration economies, IT capital complementarity, and entrepreneurship. We provide three new empirical insights into local IT labor market inequality. First, population size and supply of college-educated workers consistently drive higher regional returns to IT labor, supporting theories of labor pooling and productivity spillovers. Second, IT capital complementarity interacts with increasing returns to scale among firms that typically demand IT-skilled labor. Concentration in IT-intensive sectors is associated with higher IT wage levels, while the number of IT-intensive establishments (specifically, in IT-using services) is associated with higher wage spread. Third, IT innovation and entrepreneurship add little to the IT workers' wages, even in the most recent time period. These results hold across large MSAs within the United States, and are not peculiar to a few locations, e.g., tech hubs, or a few occupations.

Our findings suggest that supply-side policies should focus on educating students and training workers for acquiring more broad-based STEM skills, instead of narrowly focusing on the specific skills required of particularly in-demand IT jobs at the moment. Among the most effective policies were those that simply make a location more attractive to skilled workers and high IT-intensity firms. In addition, policies to encourage innovation and entrepreneurship had, at best, only a moderate effect on wage growth.

We also want to point out a few caveats in our data, which may have limited the range of analyses that this paper can run related to IT wage inequality. The wage data used in this paper contain aggregate statistics, which may not fully account for differences in the quality of skill supply across locations. For example, the same individual may be at the 25th percentile of Silicon Valley's distribution of programmer talent, but moving to El Paso or Boston might put them at the median or higher position of the local programmer wage distribution. Unfortunately, we do not have access to micro data on establishment-level employment or wages that would have allowed us to isolate the quality of skills or firm-specific productivity premiums.

Also, non-pecuniary benefits and other job characteristics are not observed in our data, another limitation that is worth noting to the reader. IT skills can be largely lucrative and rewarded by non-wage benefits, such as business income from entrepreneurial activities, vesting interest at a high-growth IT startup, or income from completing contracts with organizations that pay competitive returns to IT-related external projects. These incentives are not reflected in the OES wage statistics, and hence not studied in this paper but call for future research.

References

Akicigit, U and Sina Ates. 2019. "Ten facts on declining business dynamism and lessons from endogenous growth theory." *NBER Working Paper* .

Autor, David H. 2019. Work of the Past, Work of the Future. In *AEA Papers and Proceedings*. Vol. 109 pp. 1–32.

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.

Baqaee, David and Emmanuel Farhi. 2020. "The Darwinian Returns to Scale." *NBER Working Paper* .

Bartel, Ann, Casey Ichniowski and Kathryn Shaw. 2007. "How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills." *The Quarterly Journal of Economics* 122(4):1721–1758.

Bresnahan, Timothy F, Erik Brynjolfsson and Lorin M Hitt. 2002. "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." *The Quarterly Journal of Economics* 117(1):339–376.

Byrne, David and Carol Corrado. 2020. Accounting for innovations in consumer digital services: IT still matters. In *Measuring and Accounting for Innovation in the 21st Century*. University of Chicago Press.

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.

Cullen, Zoë and Chiara Farronato. 2020. "Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms." *Management Science* .

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.

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." *NBER Working Paper* .

Eckert, Fabian, Sharat Ganapati and Conor Walsh. 2019. "Skilled tradable services: The transformation of US high-skill labor markets." *Working Paper* .

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." *NBER Working Paper* .

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, Lawrence F Katz et al. 2020. Extending the Race between Education and Technology. In *AEA Papers and Proceedings*. Vol. 110 pp. 347–51.

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

Haltiwanger, John, Ron S Jarmin and Javier Miranda. 2013. "Who creates jobs? Small versus large versus young." *Review of Economics and Statistics* 95(2):347–361.

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 R and Frédéric Robert-Nicoud. 2020. "Tech clusters." *Journal of Economic Perspectives* 34(3):50–76.

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.

Marinoni, Astrid, John Voorheis et al. 2019. "Who Gains from Creative Destruction? Evidence from High-Quality Entrepreneurship in the United States." *Working Paper* .

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

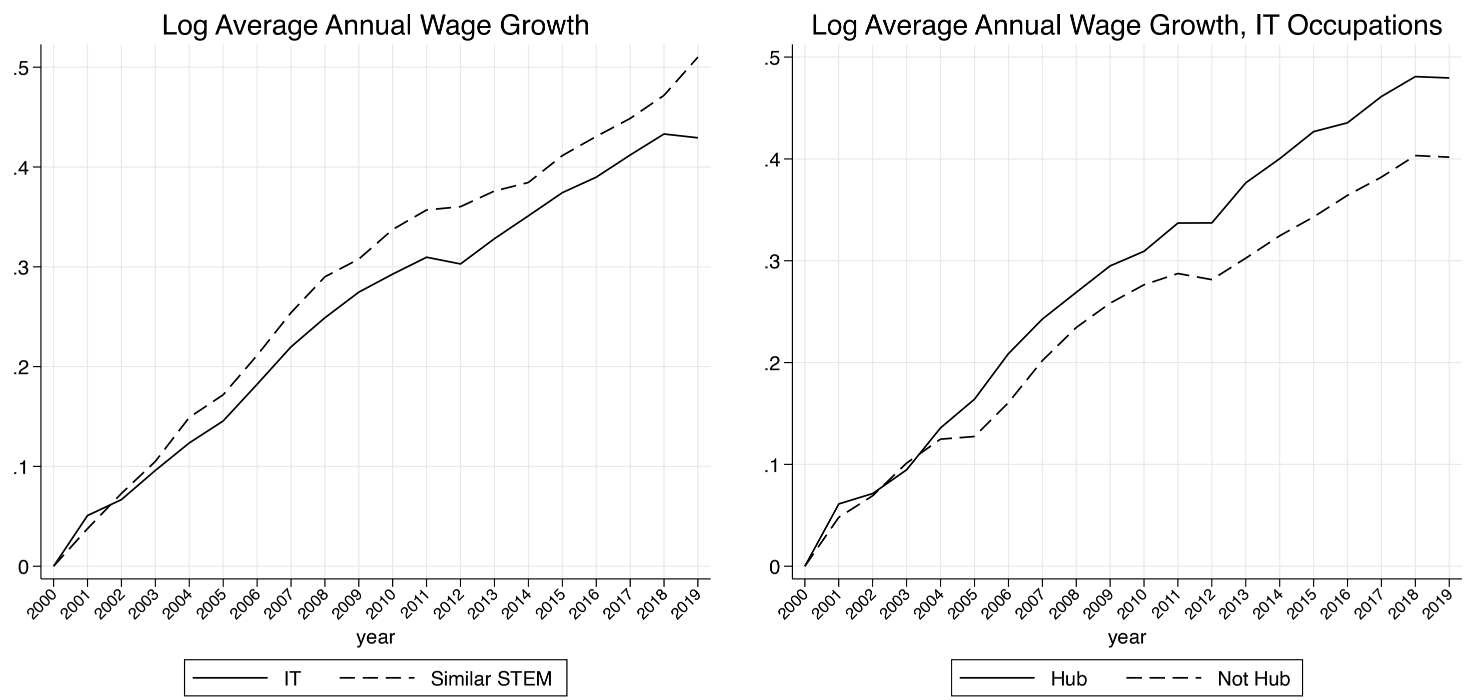
Tambe, Prasanna. 2014. "Big data investment, skills, and firm value." *Management Science* 60(6):1452–1469.

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

Tambe, Prasanna and Lorin M Hitt. 2014. "Job hopping, information technology spillovers, and productivity growth." *Management Science* 60(2):338–355.

Tambe, Prasanna, Lorin M Hitt, Daniel Rock and Erik Brynjolfsson. 2020. "Digital Capital and Superstar Firms." *NBER Working Paper* .

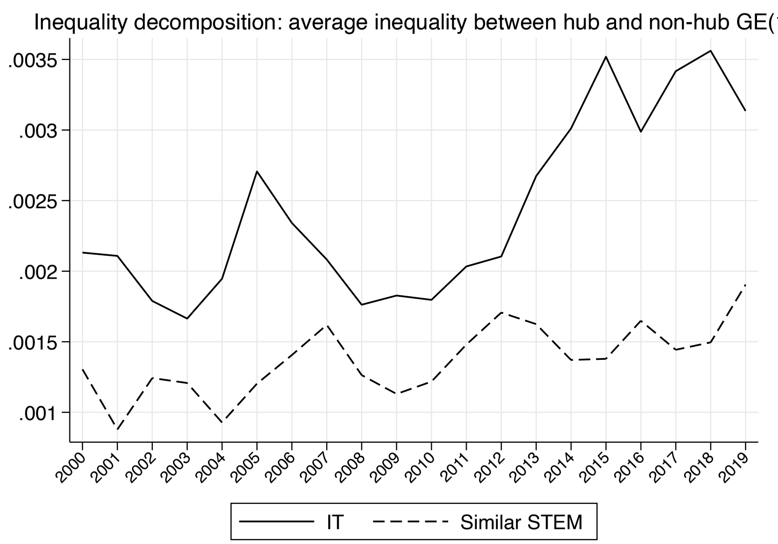
Figure 1: IT-Skilled Occupational Average Wage Growth in the United States, 2000 – 2019



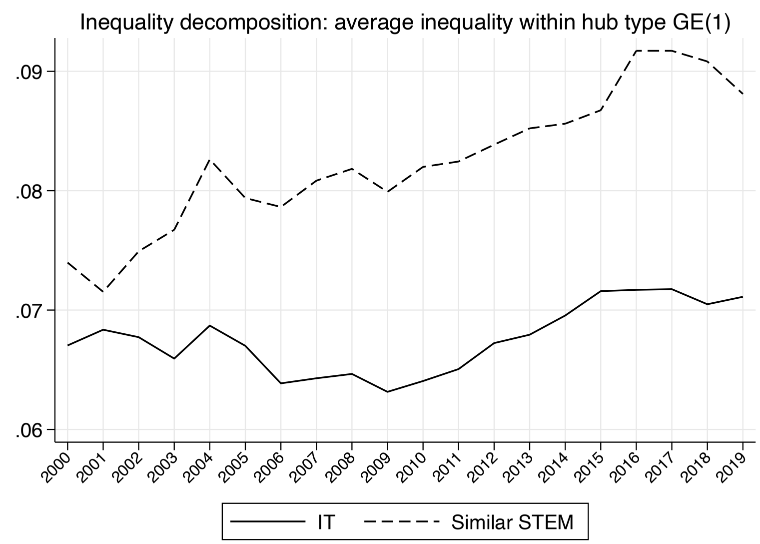
Notes: This figure shows the average year-by-year growth in the mean wage (weighted by employment size) across all 142 largest MSAs in the United States. The left panel illustrates wage growth for an average IT-skilled (including those in STEM but not strictly defined IT category) occupation, and the right panel illustrate wage growth for an average occupation only in the IT category.

Figure 2: IT-Skilled Wage Inequality Decomposition by Tech-Hub Status, 2000 – 2019

(a) GE(1) Between Tech Hubs and Non-Hub MSAs



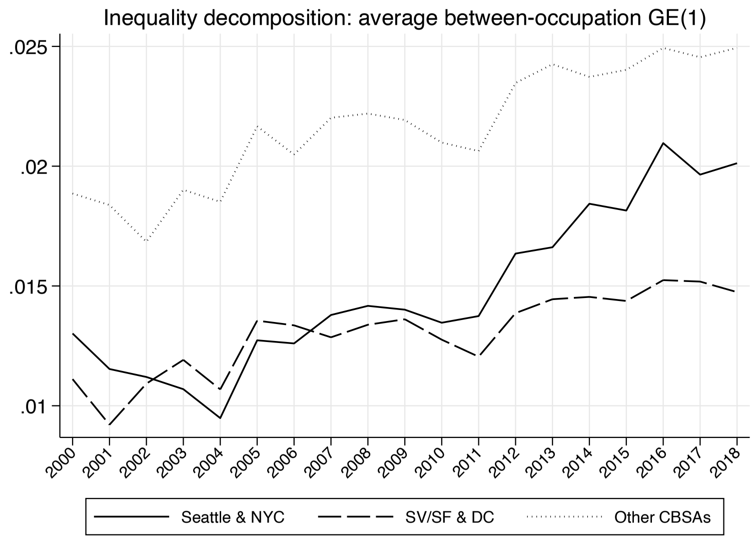
(b) GE(1) Without Tech Hubs



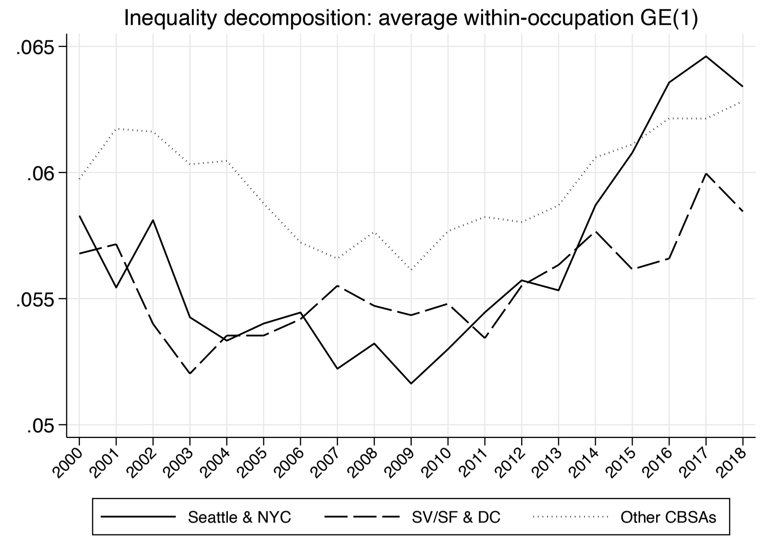
Notes: This figure presents results from a within-between decomposition of the year-by-year wage inequality index GE(1) (aka Theil's T), based on a binary group variable indicating the tech hub versus non-hub status of an MSA. Panel (a) focuses on the wage inequality between tech hubs and non-hub areas, and panel (b) focuses on the remaining inequality after accounting for the difference between tech hubs and non-hub areas. The GE(1) measure is calculated for each occupation, and each data point in the figure shows the average GE(1) over all occupations, in either IT (solid) or similarly skilled STEM (dashed) category.

Figure 3: IT Wage Inequality Decomposition by Occupation, 2000 – 2019

(a) GE(1) Between Occupations

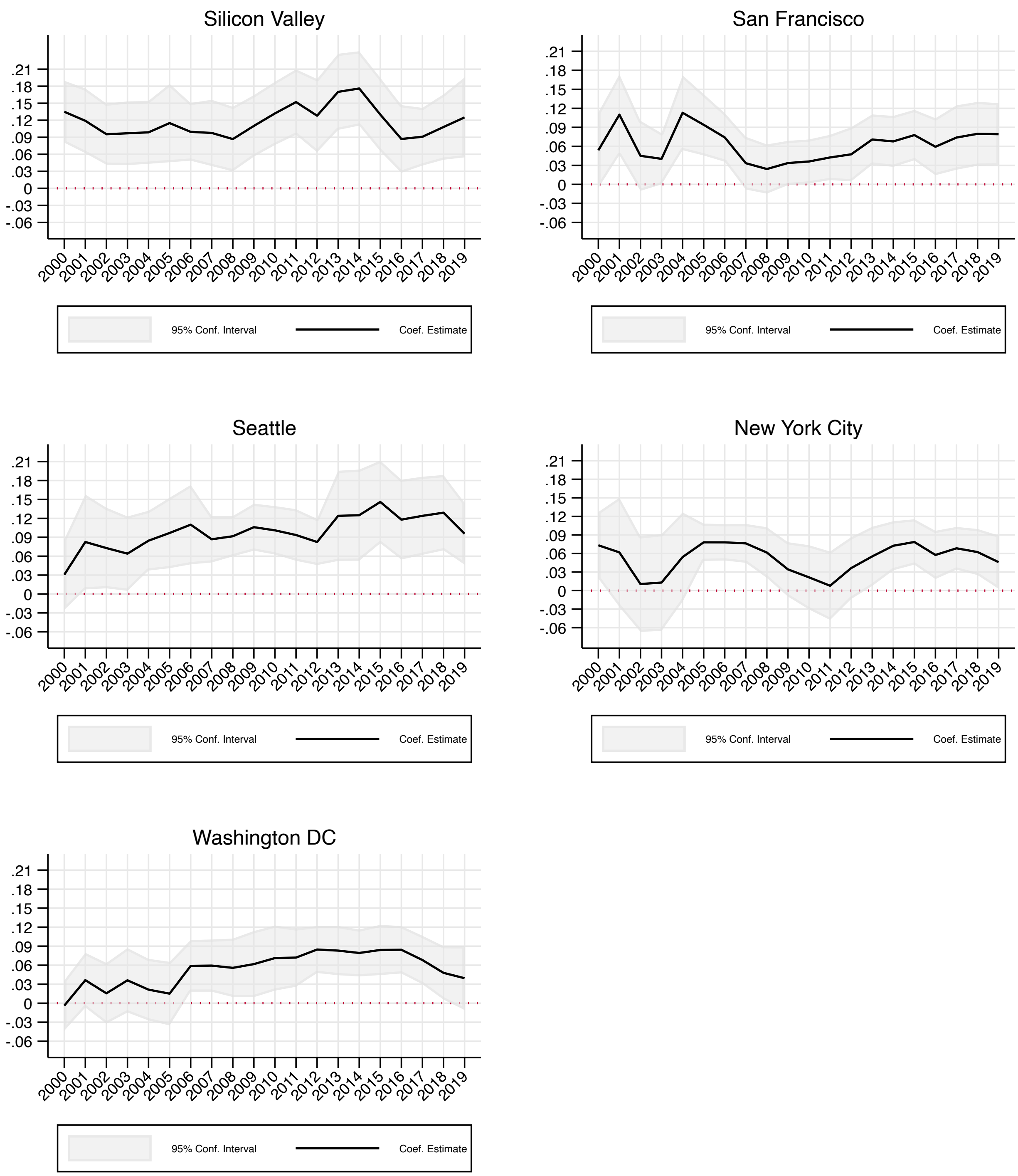


(b) Total Within-Occupation GE(1)



Notes: This figure presents results from a within-between decomposition of the year-by-year wage inequality index GE(1) (aka Theil's T), based on 6-digit occupation SOC code as the group variable. The occupations only include SOC codes in the strictly defined IT category, and do not include similarly skilled STEM occupations other than IT. Panel (a) focuses on the wage inequality between occupations, and panel (b) focuses on the remaining inequality attributed to total GE(1) within occupations. The GE(1) measure is calculated for each MSA, and each data point in the figure shows the average GE(1) over all regions, among Seattle and New York City (solid), San Francisco, Silicon Valley and Washington DC (dashed), and non-hub MSAs (dotted).

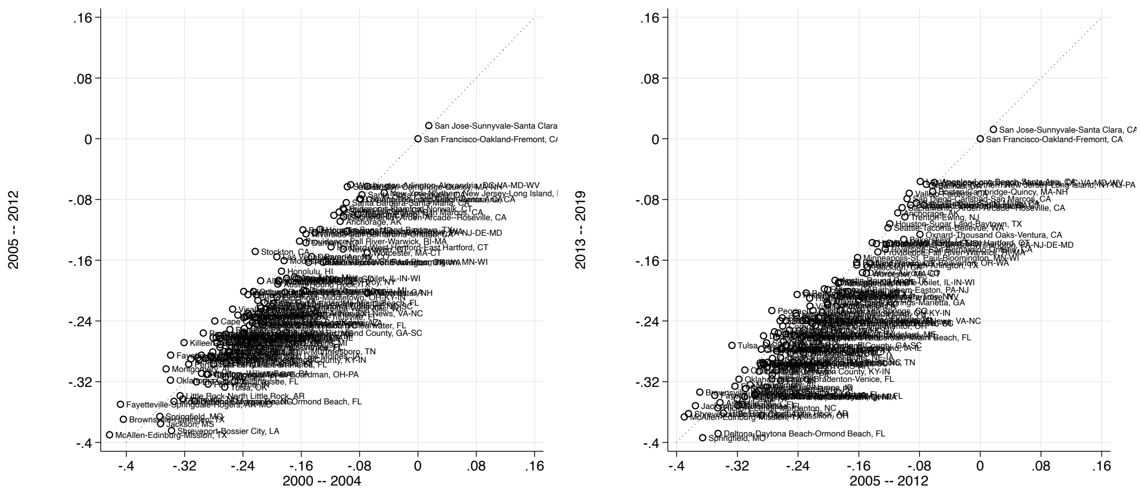
Figure 4: IT Wage Premiums in Tech Hubs, 2000 – 2019



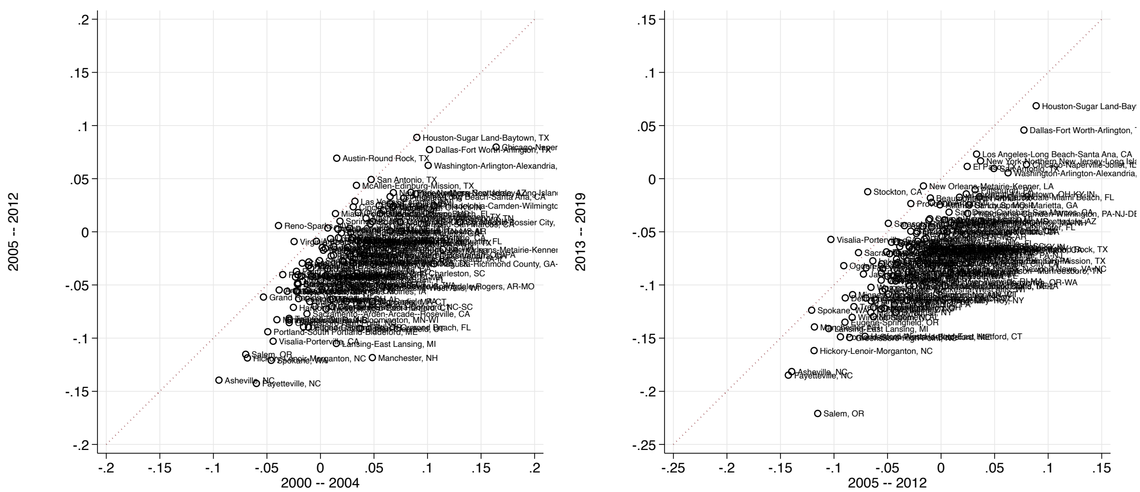
Notes: This figure plots the annual index as the estimated fixed effects on the interaction term between each tech hub MSA and the indicator for whether an occupation is in the strictly defined IT category, in equation 1. Each of the subfigures focuses on one of the five tech hubs, and plots both the coefficient estimate and the 95% confidence interval.

Figure 5: MSA Indexes Persistence, Pre-2005 v.s. 2005–2012 v.s. Post-2012

(a) Mean Wages

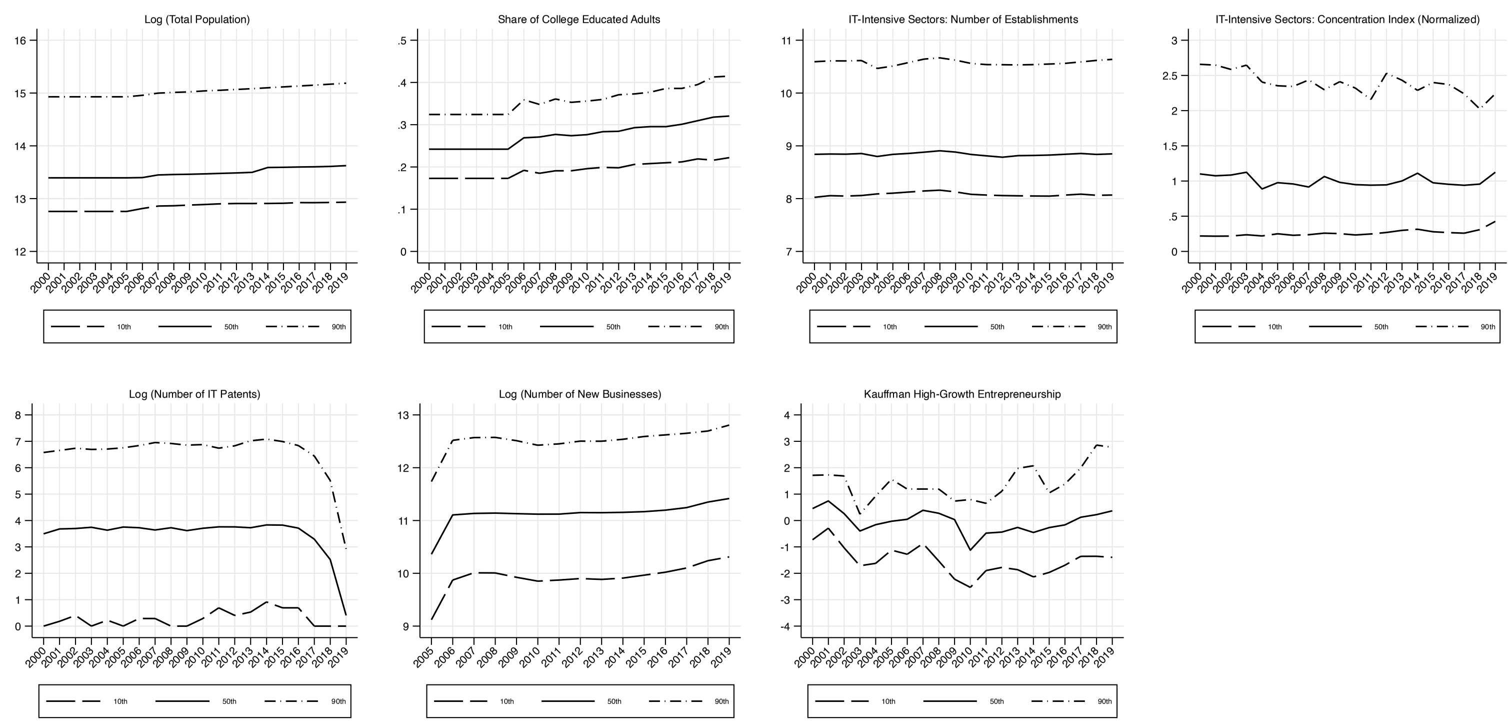


(b) 75th-to-10th-Percentile Wage Spreads



Notes: This figure shows the transition graphs for MSA indexes for the average wage and wage spread, for three time periods spanning the past two decades. The MSA indexes correspond to the fixed effects in equation 1, estimated using data on all IT-skilled occupations including STEM occupations outside the strictly defined IT category. Panel (a) focuses on the mean wage statistics, and panel (b) focuses on the 75th-to-10th-percentile wage spread. To calculate the indexes, one MSA must be omitted, and we choose San Francisco to be the omitted region in all regression specifications with MSA fixed effects.

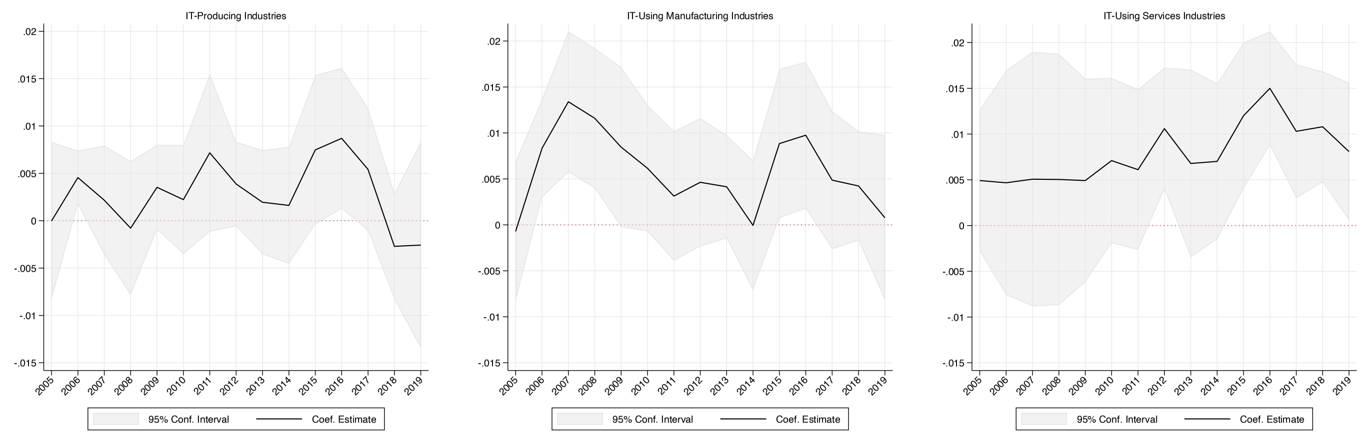
Figure 6: Summary Quantiles of MSA-Level Explanatory Variables, 2000 – 2019



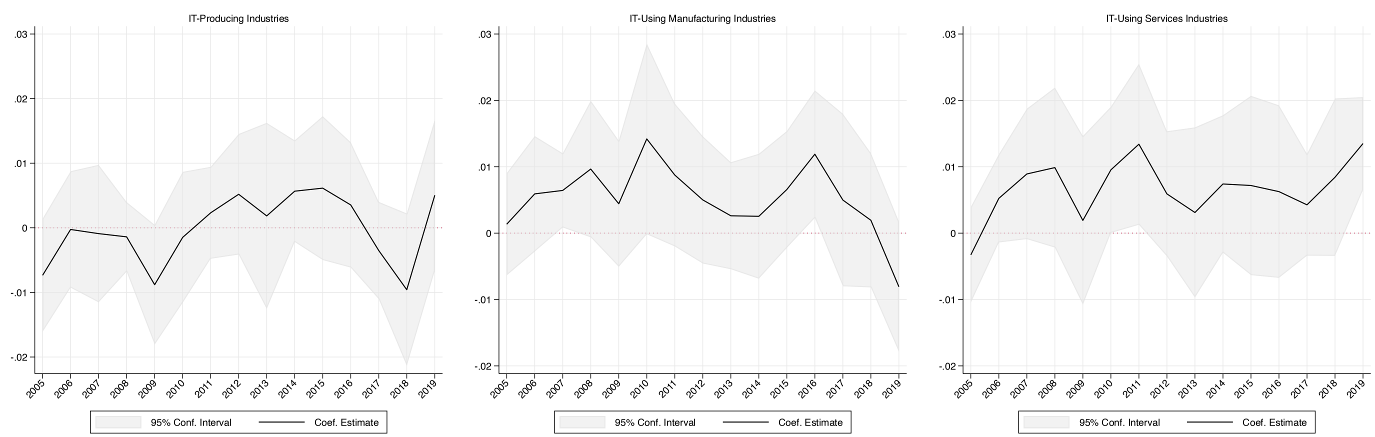
Notes: This figure shows summary statistics on explanatory variables at the MSA level, corresponding to region-based theories for determining IT labor wages outlined in Section 5.2. Each subfigure focuses on one MSA-level explanatory factor, and plots the median (solid), 10th percentile (long-dash), and 90th percentile (dot-dash) of the variable's distribution.

Figure 7: Concentration Index in IT-Intensive Sectors and IT Wage Distribution

(a) Mean Wages



(b) 75th-to-10th Percentile Wage Spreads

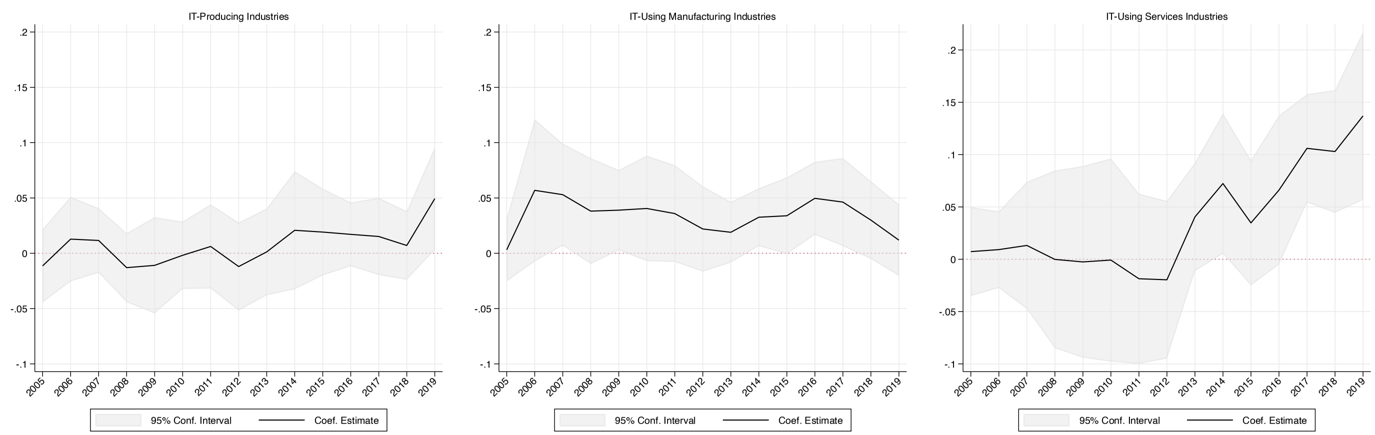


Notes: This figure plots the results from empirically estimating equations 2 and 3 on an annual basis, for all IT occupations in the 142 largest MSAs in the United States. Each sub-figure plots the estimated regression coefficients and 95% confidence intervals on the HHI concentration index for local firm establishments in the IT-producing (left), IT-using manufacturing (middle), and IT-using services (right) industries. Panel (a) focuses on the explanatory power of regressors on the average wage at the MSA level which is the outcome variable in equation 2, and panel (b) focuses on the explanatory power of regressors on the 75th-to-10th-percentile wage spread at the MSA level which is the outcome variable in equation 3. All regressions control for MSA-level demographics and supply-side characteristics, as well as occupation (6-digit SOC code) fixed effects, estimated separately for each year from 2005 to 2019. The regressions underlying panel (b) additionally control for log mean wage in the same year.

Figure 8: Establishment Counts in IT-Intensive Sectors and IT Wage Distribution

(a) Mean Wages

(b) 75th-to-10th Percentile Wage Spreads



Notes: This figure plots the results from empirically estimating equations 2 and 3 on an annual basis, for all IT occupations in the 142 largest MSAs in the United States. Each sub-figure plots the estimated regression coefficients and 95% confidence intervals on the number of local firm establishments in the IT-producing (left), IT-using manufacturing (middle), and IT-using services (right) industries. Panel (a) focuses on the explanatory power of regressors on the average wage at the MSA level which is the outcome variable in equation 2, and panel (b) focuses on the explanatory power of regressors on the 75th-to-10th-percentile wage spread at the MSA level which is the outcome variable in equation 3. All regressions control for MSA-level demographics and supply-side characteristics, as well as occupation (6-digit SOC code) fixed effects, estimated separately for each year from 2005 to 2019. The regressions underlying panel (b) additionally control for log mean wage in the same year.

Table 1: Examples of IT-Skilled Wage Statistics by Metroplitan Statistical Area (MSA) and Occupation

	San Francisco, CA				**Indianapolis, IN**				**Little Rock, AR**			
	75th	Mean	10th	Rank	75th	Mean	10th	Rank	75th	Mean	10th	Rank
Computer Scientists, Research												
2000	112230	89420	52620	2	65270	53450	36750	51	.	.	.	.
2009	136380	119970	71260	8	118890	101710	65340	29	.	.	.	.
2018	176780	140660	66270	6	140090	117260	69100	25	.	.	.	.
Database Administrators												
2000	84380	62360	32850	11	69690	53540	30780	54	.	.	.	.
2009	112330	88330	49720	2	81630	65170	34130	90	74980	62450	39340	102
2018	139110	107660	58620	5	98440	78730	42880	102	94170	79150	51000	101
Computer User Support Specialists												
2000	65590	52830	31900	3	44140	39800	25470	46	37600	31640	20100	120
2009	72230	60000	35690	4	51930	42680	26170	91	46570	39000	25020	126
2018	88870	73600	43420	2	60180	50090	31880	78	54540	46690	31210	113
Engineering Managers												
2000	130230	102150	63480	2	86780	74930	51580	103	79940	66380	45630	139
2009	171488	146950	93910	3	104030	92580	67710	134	128220	106320	74110	92
2018	212125	184130	102850	5	144710	121510	75090	122	137880	118030	78160	129
Statisticians												
2000	82180	73110	57630	1	.	.	.	.	46460	40790	25680	51
2009	111970	92070	53840	3	78500	60760	33290	53	44900	40560	29690	75
2018	147420	118570	69380	2	93920	70750	39430	74	74570	58690	36680	87
Surveying and Mapping Technicians												
2000	70020	63460	39910	1	34740	30430	22500	69	28830	26400	18620	102
2009	74300	61500	44890	1	43430	38080	26530	74	40690	33890	18760	112
2018	93140	72190	37620	3	53070	43230	28850	89	46910	38860	24970	113

Notes: This table showcases a few examples in our final wage data set. We focus on three locations with IT wages among the highest (San Francisco CA), close-to-middle (Indianapolis IN), and lowest (Little Rock AR) of overall 142 largest MSAs. We show the wage statistics for three strictly defined IT occupations – information and computer research scientists (top wage), database administrators (middle wage), and user support specialists (bottom wage) – as well as three non-IT STEM occupations typically performing IT-related tasks – engineering managers (top wage), statisticians (middle wage), and surveying and mapping technicians (bottom wage). For each MSA and occupation, we list the 75th percentile, the mean, and the 10th percentile of the wage distribution, as well as the relative MSA ranking of the average wage. We show these wage statistics for every 9 years spanning the last two decades, i.e., in 2000, 2009, and 2018 respectively.

Table 2: Explanatory Local Variables and Log Mean IT Wages, 2013 – 2019

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.022***	0.030***	0.108***	0.106***	0.018***	0.017***	0.016**	0.094***
	(0.005)	(0.005)	(0.028)	(0.028)	(0.006)	(0.006)	(0.006)	(0.025)
Share of College-Educated Adults	0.146	0.146	0.261*	0.250*	0.107	0.241	0.218	0.308*
	(0.126)	(0.119)	(0.134)	(0.126)	(0.121)	(0.154)	(0.151)	(0.153)
IT-Intensive Sectors: Employment HHI		0.011**		0.009**				0.008*
		(0.004)		(0.004)				(0.004)
IT-Intensive Sectors: Number of Establishments			-0.082***	-0.074***				-0.070***
			(0.027)	(0.026)				(0.023)
Log (Number of IT Patents)					0.004		0.002	0.002
					(0.002)		(0.002)	(0.002)
Log (Number of New Businesses)						0.017***	0.017***	0.015***
						(0.006)	(0.006)	(0.006)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	11278	11278	11278	11278	11278	11278	11278	11278
R^2	0.817	0.818	0.819	0.820	0.818	0.820	0.820	0.822
Mean Outcome	11.254	11.254	11.254	11.254	11.254	11.254	11.254	11.254

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2013 to 2019. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table 3: Explanatory Local Variables and Log Mean IT Wages, 2005 – 2012

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.018***	0.025***	0.104***	0.101***	0.012*	0.015**	0.013*	0.086***
	(0.005)	(0.006)	(0.026)	(0.026)	(0.007)	(0.006)	(0.007)	(0.028)
Share of College-Educated Adults	0.161	0.143	0.312**	0.283**	0.113	0.240*	0.213*	0.319**
	(0.125)	(0.112)	(0.140)	(0.124)	(0.115)	(0.132)	(0.119)	(0.140)
IT-Intensive Sectors: Employment HHI		0.010*		0.008				0.007
		(0.006)		(0.006)				(0.006)
IT-Intensive Sectors: Number of Establishments			-0.083***	-0.074***				-0.065**
			(0.025)	(0.025)				(0.027)
Log (Number of IT Patents)					0.004		0.002	0.002
					(0.002)		(0.002)	(0.002)
Log (Number of New Businesses)						0.019***	0.018***	0.017***
						(0.005)	(0.006)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10767	10767	10767	10767	10767	10767	10767	10767
R^2	0.805	0.806	0.807	0.808	0.806	0.809	0.809	0.811
Mean Outcome	11.124	11.124	11.124	11.124	11.124	11.124	11.124	11.124

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2005 to 2012. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table 4: Explanatory Local Variables and Log Mean IT Wages, 2000 – 2004

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.020***	0.016*	-0.024**	-0.025**	0.012	0.020***	0.012	-0.025*
	(0.006)	(0.008)	(0.011)	(0.011)	(0.008)	(0.006)	(0.009)	(0.014)
Share of College-Educated Adults	0.316*	0.304*	0.205	0.205	0.276*	0.308**	0.266*	0.183
	(0.163)	(0.166)	(0.155)	(0.153)	(0.157)	(0.147)	(0.143)	(0.133)
IT-Intensive Sectors: Employment HHI		-0.005		0.007				0.007
		(0.006)		(0.006)				(0.006)
IT-Intensive Sectors: Number of Establishments			0.046***	0.053***				0.048***
			(0.010)	(0.012)				(0.014)
Log (Number of IT Patents)					0.006*		0.006*	0.004
					(0.003)		(0.003)	(0.003)
Kauffman High-Growth Entrepreneurship						0.002	0.002	0.002
						(0.005)	(0.005)	(0.004)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	5819	5819	5819	5819	5819	5819	5819	5819
R^2	0.755	0.755	0.760	0.760	0.757	0.755	0.757	0.761
Mean Outcome	10.947	10.947	10.947	10.947	10.947	10.947	10.947	10.947

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2000 to 2004. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table 5: Explanatory Local Variables and IT Wage Spreads, 2013 – 2019

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	-0.064	-0.066	-0.056	-0.058	-0.066	-0.084	-0.084	-0.077
	(0.046)	(0.047)	(0.046)	(0.046)	(0.047)	(0.057)	(0.057)	(0.058)
Log (Total Population)	0.024***	0.026***	-0.054*	-0.055*	0.020***	0.018***	0.018***	-0.067**
	(0.005)	(0.006)	(0.029)	(0.029)	(0.005)	(0.005)	(0.006)	(0.030)
Share of College-Educated Adults	-0.065	-0.064	-0.170	-0.177	-0.094	0.052	0.046	-0.067
	(0.097)	(0.098)	(0.112)	(0.114)	(0.099)	(0.094)	(0.100)	(0.116)
IT-Intensive Sectors: Employment HHI		0.004		0.005				0.005
		(0.004)		(0.004)				(0.004)
IT-Intensive Sectors: Number of Establishments			0.074***	0.079***				0.084***
			(0.027)	(0.027)				(0.029)
Log (Number of IT Patents)					0.003		0.000	0.000
					(0.002)		(0.002)	(0.002)
Log (Number of New Businesses)						0.021***	0.021***	0.021***
						(0.005)	(0.005)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	11278	11278	11278	11278	11278	11278	11278	11278
R^2	0.127	0.127	0.130	0.130	0.127	0.136	0.136	0.140
Mean Outcome	0.722	0.722	0.722	0.722	0.722	0.722	0.722	0.722

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2013 to 2019. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table 6: Explanatory Local Variables and IT Wage Spreads, 2005 – 2012

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	-0.067	-0.071	-0.068	-0.070	-0.067	-0.075	-0.075	-0.077
	(0.052)	(0.051)	(0.051)	(0.051)	(0.052)	(0.053)	(0.053)	(0.052)
Log (Total Population)	0.019***	0.024***	0.020	0.017	0.020***	0.018***	0.020***	0.014
	(0.004)	(0.005)	(0.028)	(0.028)	(0.005)	(0.004)	(0.004)	(0.029)
Share of College-Educated Adults	0.144	0.134	0.145	0.123	0.147	0.178*	0.195*	0.170
	(0.090)	(0.089)	(0.105)	(0.107)	(0.092)	(0.089)	(0.100)	(0.102)
IT-Intensive Sectors: Employment HHI		0.006		0.006				0.006
		(0.004)		(0.005)				(0.005)
IT-Intensive Sectors: Number of Establishments			-0.000	0.006				0.010
			(0.027)	(0.027)				(0.026)
Log (Number of IT Patents)					-0.000		-0.001	-0.001
					(0.002)		(0.002)	(0.002)
Log (Number of New Businesses)						0.008*	0.008*	0.008*
						(0.004)	(0.004)	(0.004)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10767	10767	10767	10767	10767	10767	10767	10767
R^2	0.146	0.147	0.146	0.147	0.146	0.147	0.148	0.148
Mean Outcome	0.698	0.698	0.698	0.698	0.698	0.698	0.698	0.698

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2005 to 2012. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table 7: Explanatory Local Variables and IT Wage Spreads, 2000 – 2004

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	0.142***	0.144***	0.144***	0.141***	0.144***	0.142**	0.145**	0.143**
	(0.039)	(0.039)	(0.040)	(0.040)	(0.039)	(0.054)	(0.054)	(0.056)
Log (Total Population)	0.013**	0.021***	0.017**	0.016**	0.015***	0.013**	0.015***	0.016
	(0.005)	(0.005)	(0.008)	(0.007)	(0.006)	(0.006)	(0.006)	(0.010)
Share of College-Educated Adults	-0.050	-0.027	-0.040	-0.039	-0.037	-0.040	-0.026	-0.018
	(0.120)	(0.116)	(0.120)	(0.118)	(0.119)	(0.127)	(0.123)	(0.127)
IT-Intensive Sectors: Employment HHI		0.011**		0.012**				0.013*
		(0.004)		(0.005)				(0.007)
IT-Intensive Sectors: Number of Establishments			-0.004	0.007				0.010
			(0.008)	(0.008)				(0.009)
Log (Number of IT Patents)					-0.002		-0.002	-0.002
					(0.002)		(0.002)	(0.002)
Kauffman High-Growth Entrepreneurship						-0.002	-0.002	-0.003
						(0.005)	(0.005)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	5819	5819	5819	5819	5819	5819	5819	5819
R^2	0.204	0.205	0.204	0.206	0.204	0.204	0.204	0.206
Mean Outcome	0.696	0.696	0.696	0.696	0.696	0.696	0.696	0.696

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within the 142 largest MSAs in the United States, in the time period from 2000 to 2004. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

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.

Business Formation and Entrepreneurship. For new business formation and entrepreneurial activities, we use two sets of aggregate data – Census Business Formation and Kauffman Indicators[21]. These data are available at the state level, for all years from 1999 to 2017.

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.

O*NET Database. We use occupational task composition and skill requirement data in the O*NET Database to identify STEM occupations that perform similar job tasks to IT occupations. Each occupation is identified with a 6-digit SOC code, and the data contains all the indirect work activities (IWA) associated with an occupation. We narrow down the set

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

of STEM occupations to those sharing 20% of IWAs with at least one of the IT occupations in O*NET's July 2014 release (v19.0).

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.

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 Decomposing Wage Inequality from Summary Statistics

The BLS Occupational Employment Statistics (OES) data contains, for each occupation (defined by the Standard Occupation Classification codes) and MSA, 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 full 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.

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 MSA (for the largest 142 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

$$E[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \tag{4}$$

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

$$y_\tau = \exp\left(\mu + \sqrt{2\sigma^2} \cdot erf^{-1}(2\tau - 1)\right) \tag{5}$$

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 4 and 5, 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 3 reports the between-region component of wage inequality using GE(1), aka Theil's T as the measure for inequality. Other inequality indexes such as GE(0) and total variance yield very similar trends[22].

A.3 Occupation Indexes for IT Wages

We describe patterns in the occupation-specific wage indexes from 2000 to 2019, for the set of all 6-digit SOC codes that represent IT occupations specifically. We estimate Equation 1 on the wage summary statistics at the occupation and MSA levels of IT and a subset of STEM occupations that share 20% or more of the indirect work activities of an IT occupation. The

[22] We do not include these graphs in the paper, but they are available upon request by emailing the authors.

occupation-specific wage indexes I^{occ} are estimated for all the 6-digit SOC codes with data available for at least 10% of the largest 142 MSAs in the sample.

The OCC classification codes are modified a few times during the time period of 2000 to 2019. A different code system applies to 2000 – 2009, 2010 – 2018, and 2019[23]. Within the same code system, minor adjustments were made over time, which changed the number of 6-digit OCC codes pertaining to IT occupations from year to year. For example, from 2010 to 2011 there were 9 IT occupations, among which is a combined category 15-1179 later split up into standalone codes such as 15-1122 Information Security Analysts, 15-1134 Web Developers, and 15-1143 Computer Network Architects after 2011. The combined category 15-1150 Computer Support Specialists was also split into standalone codes 15-1151 Computer User Support Specialists and 15-1152 Computer Network Support Specialists after 2011.

Despite these changes to occupational code classifications across years, IT occupations in each year can be broadly categorized as one of the five distinct types: research scientists (2010 SOC code 15-1111), developers (2010 SOC code 15-113X)[24], computer information analysts (2010 SOC code 15-112X), database and network administrators (2010 SOC code 15-114X), and computer support specialists (2010 SOC code 15-115X). Figure A.3 top-left panel – A.3 bottom-left panel illustrate patterns in the relative magnitudes of occupational wage indexes between research scientists and the rest of IT occupation types. Research scientists have been consistently the highest paying profession among IT jobs, and also require a higher level of education (master's degree) compared to the rest of the IT occupations (which require at most a bachelor's degree). The equivalent index for research scientists is 0 (and standard errors are undefined), as we needed to omit one occupation category each year, and chose research scientists to be the omitted category in order to estimate the fixed effects in Equation 1.

According to Figure A.3 top-left panel, whereas developer jobs were only paid about 5% less than research scientists in the early 2000s, the occupational wage indexes have declined by 20% relative to research scientists between then and 2019, and currently wages are 25%

[23]The code system for occupational wage data is called the Standard Occupation Classification (SOC) System. There has been three different major revisions over the last two decades: the 2000 version which defines occupations from 2000 to 2009, the 2010 version which defines occupations from 2010 to 2018, and the 2018 version which defines occupations from 2019 onward.

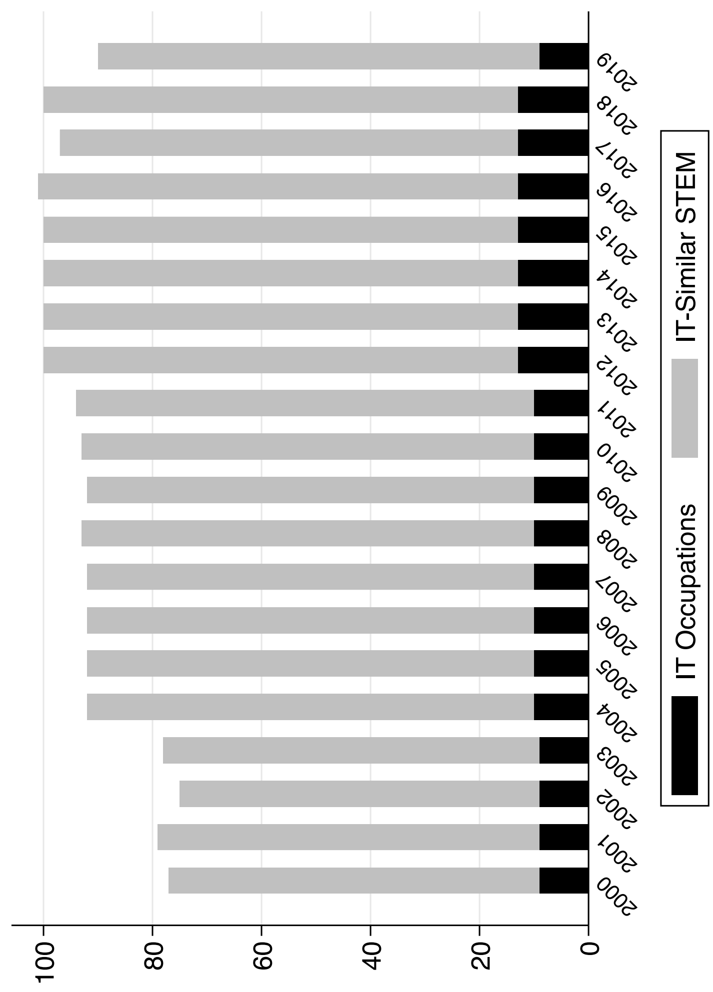
[24]Including programmers, software engineers and web developers

lower for developers compared to research scientists. On the other hand, Figure A.3 bottom-right panel shows that computer support specialists have always been the lowest paying IT occupation, where wages are between 50 – 80% lower than research scientists, and the gap widened slowly from one year to the next, except for the time period around the financial crisis which may have been followed by a multi-year recovery from declining wages across all types of occupational wages (including those of research scientists).

Among the middle-wage earning IT occupations, both database and network administrators (Figure A.3 top-right panel and computer information analysts (Figure A.3 bottom-left panel experience consistently declining relative wage indexes since 2000, and until the 2008 Financial Crisis the wage levels have declined by 12 – 15% relative to research scientists. However, the crisis period seems to have acted as an equalizing force for these middle-wage occupations, where they caught up to their 2000 levels of relative wages. After 2012, the wage gaps seem to remain stable and did not continue to shrink between these occupations and research scientist.

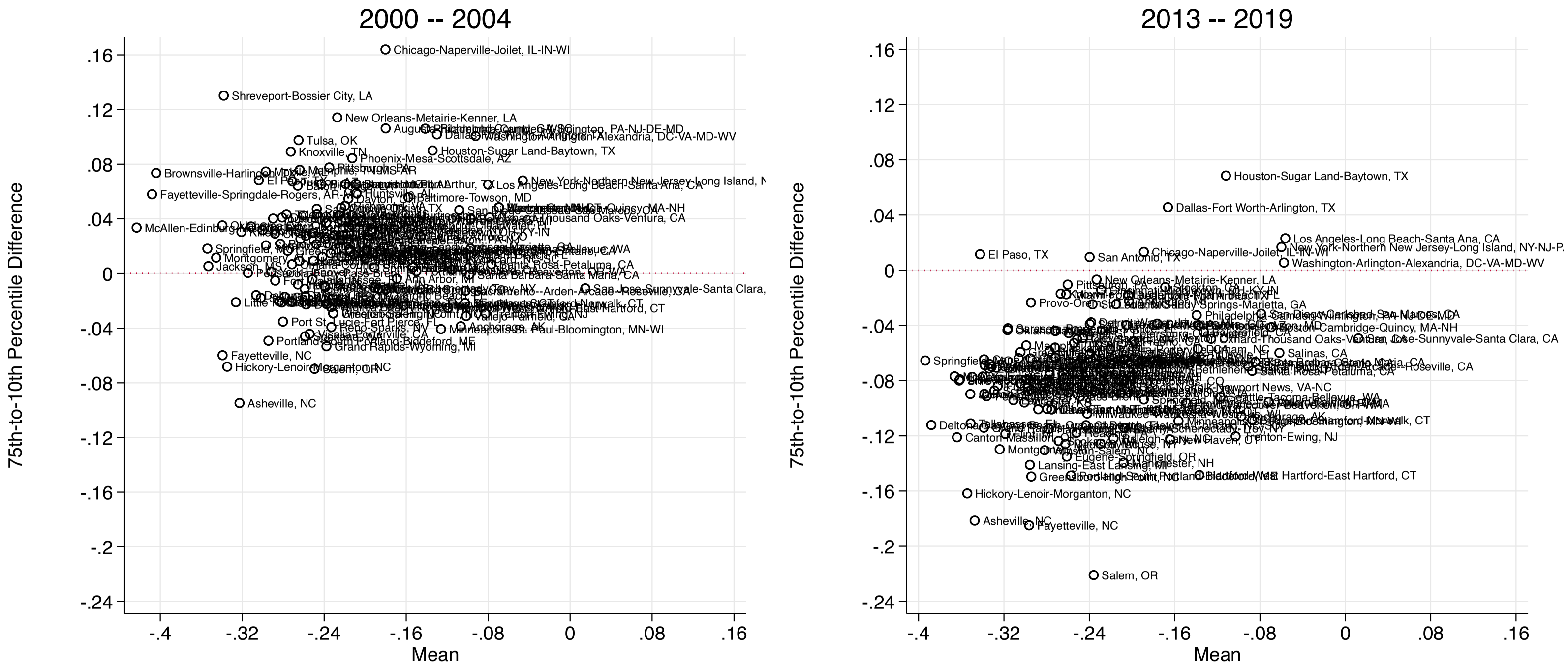
A.4 Appendix Figures and Tables

Figure A.1: Number of IT-Skilled Occupations by Year, 2000 – 2019



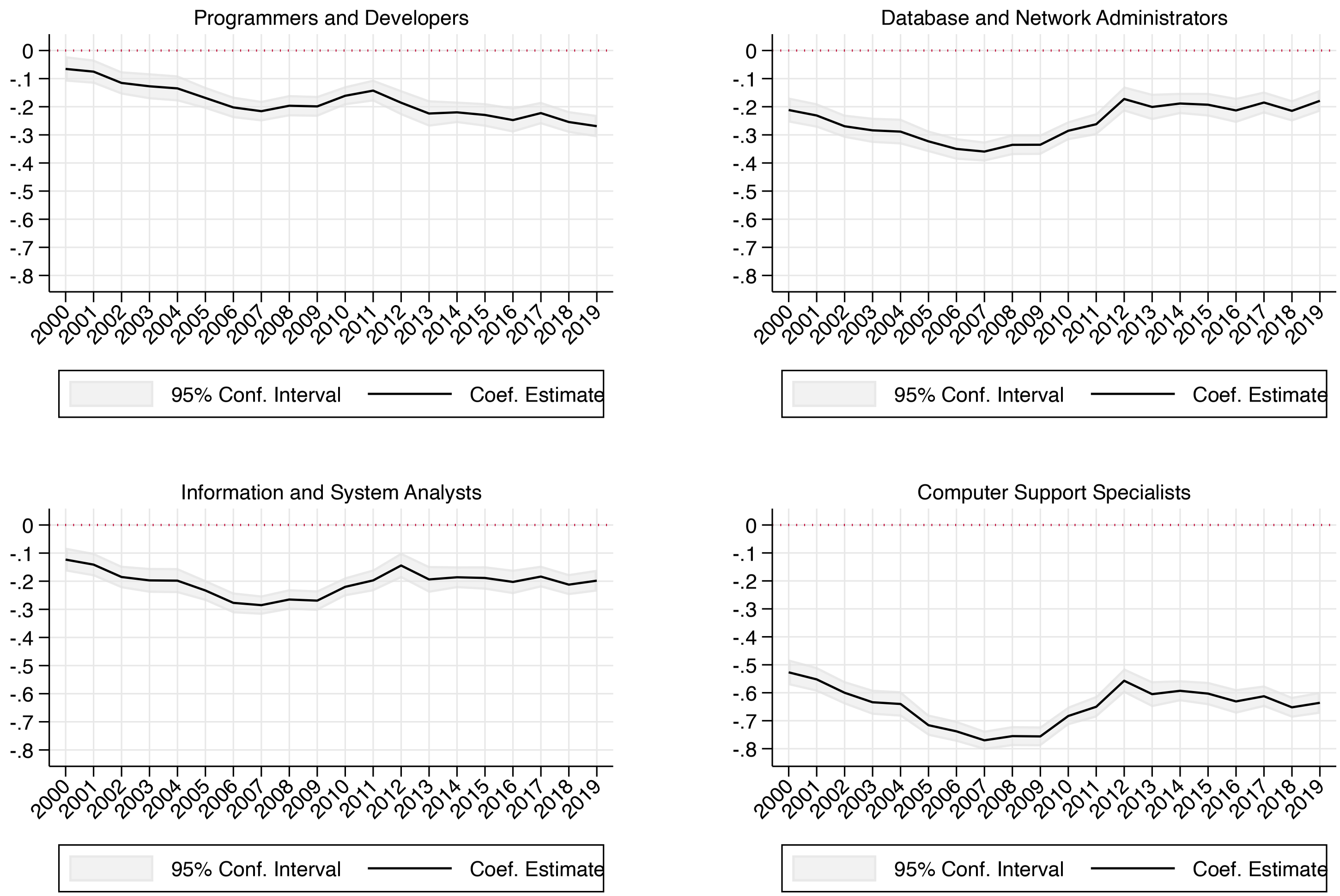
Notes: This figure shows the number of occupations in strictly defined IT category (black) and non-IT STEM category sharing 20% or more IT indirect work activities (shaded), from 2000 to 2019 on an annual basis.

Figure A.2: MSA Indexes for IT-Skilled Wages: Mean v.s. 75th-to-10th-Percentile Spread



Notes: This figure shows scatter plots of the relationship between the MSA indexes for the mean and the 75th-to-10th-percentile spread in IT-skilled labor wages. The left panel focuses on data from 2000 to 2004, and the right panel focuses on data from 2013 to 2019. The indexes are estimated on all IT-skilled occupations, both those strictly defined in the IT category and those in the non-IT STEM category requiring 20% or more IT-related indirect work activities (IWAs). To calculate the indexes, one MSA must be omitted, and we choose San Francisco to be the omitted region in all regression specifications with MSA fixed effects.

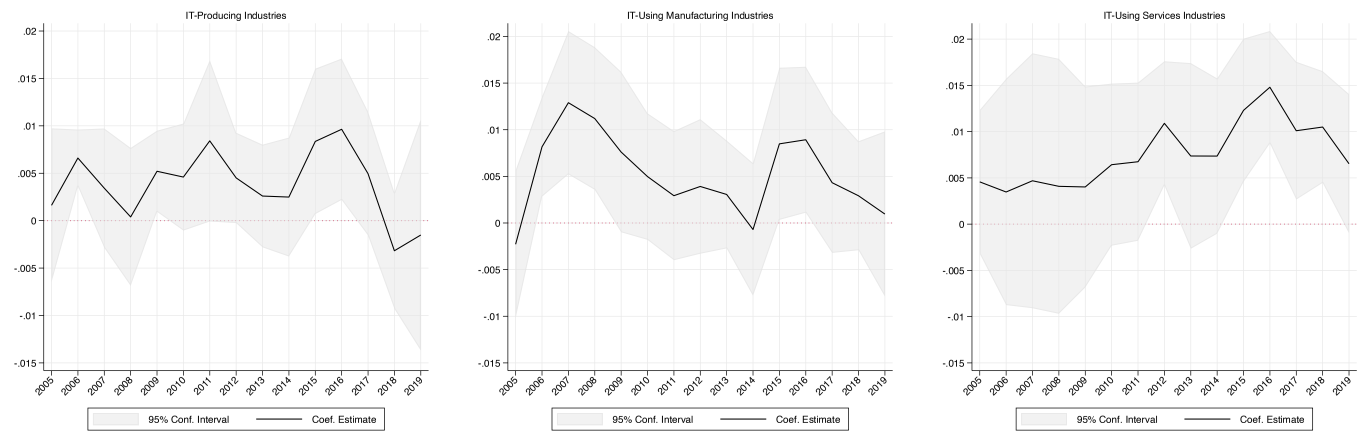
Figure A.3: IT Occupation Indexes for Mean Wages, 2000 – 2019



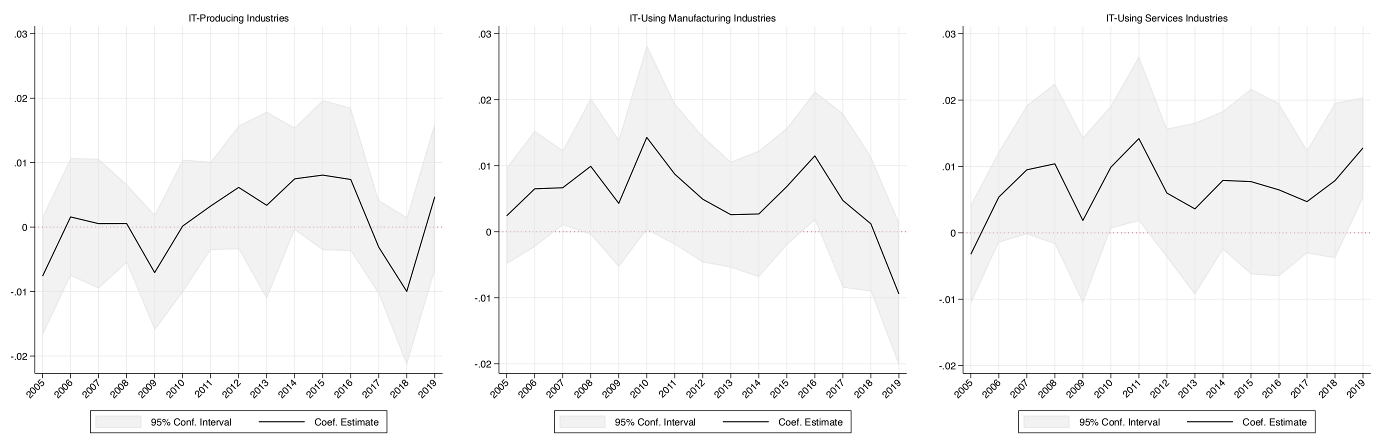
Notes: This figure plots the occupation indexes, for each of the four IT occupation types – programmers and developers (top left), database and network administrators (top right), information and system analysts (bottom left), and computer support specialists (bottom right) – estimated from equation 1 using the mean wage statistic as the outcome variable. Each subfigure plots the averages of occupation fixed-effect estimates and 95% confidence intervals across 6-digit SOC codes, on an annual basis from 2000 to 2019. All the indexes reflect relative percentage differences from the omitted category (i.e., information and research scientists).

Figure A.4: Concentration Index in IT-Intensive Sectors and IT Wages Outside Tech Hubs

(a) Mean Wages



(b) 75th-to-10th Percentile Wage Spreads

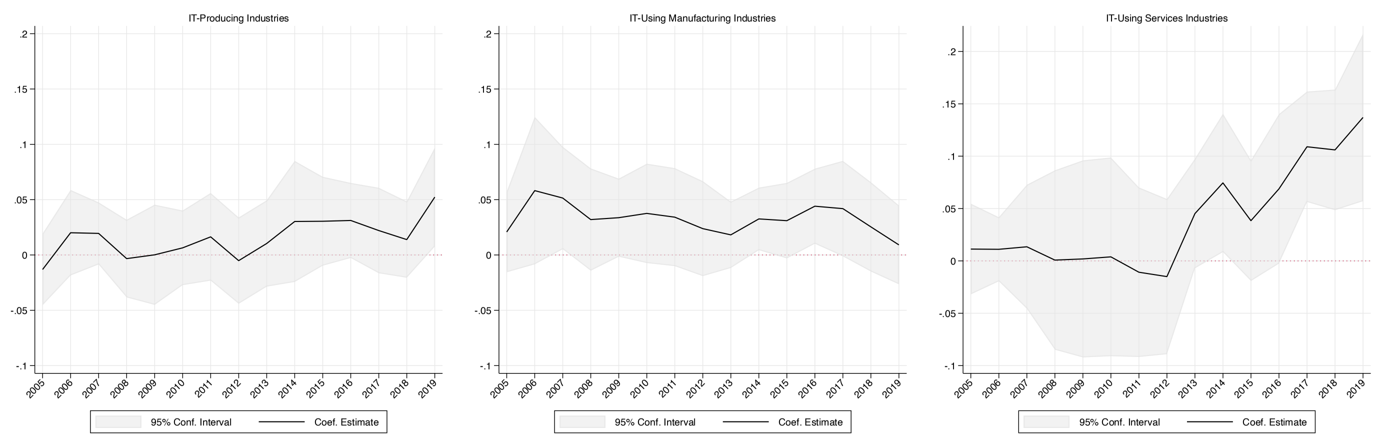


Notes: This figure plots the results from empirically estimating equations 2 and 3 on an annual basis, for all IT occupations in 137 large MSAs excluding the tech hubs. Each sub-figure plots the estimated regression coefficients and 95% confidence intervals on the HHI concentration index for local firm establishments in the IT-producing (left), IT-using manufacturing (middle), and IT-using services (right) industries. Panel (a) focuses on the explanatory power of regressors on the average wage at the MSA level which is the outcome variable in equation 2, and panel (b) focuses on the explanatory power of regressors on the 75th-to-10th-percentile wage spread at the MSA level which is the outcome variable in equation 3.

Figure A.5: Establishment Counts in IT-Intensive Sectors and IT Wages Outside Tech Hubs

(a) Mean Wages

(b) 75th-to-10th Percentile Wage Spreads



Notes: This figure plots the results from empirically estimating equations 2 and 3 on an annual basis, for all IT occupations in 137 large MSAs excluding the tech hubs. Each sub-figure plots the estimated regression coefficients and 95% confidence intervals on the number of local firm establishments in the IT-producing (left), IT-using manufacturing (middle), and IT-using services (right) industries. Panel (a) focuses on the explanatory power of regressors on the average wage at the MSA level which is the outcome variable in equation 2, and panel (b) focuses on the explanatory power of regressors on the 75th-to-10th-percentile wage spread at the MSA level which is the outcome variable in equation 3.

Table A.1: Explanatory Local Variables and Log Mean IT Wages (Excluding Tech Hubs), 2013 – 2019

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.022***	0.030***	0.111***	0.110***	0.020***	0.017***	0.016**	0.099***
	(0.005)	(0.006)	(0.028)	(0.027)	(0.007)	(0.006)	(0.007)	(0.025)
Share of College-Educated Adults	0.162	0.161	0.284**	0.272**	0.146	0.245	0.239	0.335**
	(0.127)	(0.120)	(0.134)	(0.126)	(0.124)	(0.163)	(0.162)	(0.165)
IT-Intensive Sectors: Employment HHI		0.011**		0.009**				0.008*
		(0.004)		(0.004)				(0.004)
IT-Intensive Sectors: Number of Establishments			-0.086***	-0.078***				-0.072***
			(0.026)	(0.025)				(0.022)
Log (Number of IT Patents)					0.001		0.000	0.000
					(0.003)		(0.002)	(0.002)
Log (Number of New Businesses)						0.015**	0.015**	0.014**
						(0.006)	(0.006)	(0.006)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10848	10848	10848	10848	10848	10848	10848	10848
R^2	0.807	0.808	0.809	0.810	0.807	0.809	0.809	0.812
Mean Outcome	11.242	11.242	11.242	11.242	11.242	11.242	11.242	11.242

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2013 to 2019. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table A.2: Explanatory Local Variables and Log Mean IT Wages (Excluding Tech Hubs), 2005 – 2012

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.018***	0.025***	0.109***	0.106***	0.015**	0.015**	0.014*	0.093***
	(0.005)	(0.006)	(0.026)	(0.025)	(0.007)	(0.006)	(0.008)	(0.028)
Share of College-Educated Adults	0.158	0.140	0.321**	0.292**	0.132	0.228	0.215*	0.334**
	(0.124)	(0.111)	(0.136)	(0.120)	(0.115)	(0.140)	(0.126)	(0.145)
IT-Intensive Sectors: Employment HHI		0.010*		0.007				0.007
		(0.006)		(0.006)				(0.006)
IT-Intensive Sectors: Number of Establishments			-0.087***	-0.079***				-0.070***
			(0.025)	(0.024)				(0.026)
Log (Number of IT Patents)					0.002		0.001	0.000
					(0.003)		(0.002)	(0.002)
Log (Number of New Businesses)						0.016***	0.016***	0.014**
						(0.005)	(0.006)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10355	10355	10355	10355	10355	10355	10355	10355
R^2	0.795	0.796	0.797	0.798	0.795	0.798	0.798	0.800
Mean Outcome	11.114	11.114	11.114	11.114	11.114	11.114	11.114	11.114

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2005 to 2012. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table A.3: Explanatory Local Variables and Log Mean IT Wages (Excluding Tech Hubs), 2000 – 2004

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Total Population)	0.020***	0.016*	-0.025**	-0.026**	0.016*	0.020***	0.016	-0.026*
	(0.006)	(0.009)	(0.011)	(0.011)	(0.009)	(0.006)	(0.011)	(0.014)
Share of College-Educated Adults	0.448***	0.434***	0.332**	0.335**	0.411**	0.445***	0.406**	0.337**
	(0.152)	(0.156)	(0.142)	(0.139)	(0.160)	(0.135)	(0.151)	(0.126)
IT-Intensive Sectors: Employment HHI		-0.005		0.009				0.009
		(0.006)		(0.006)				(0.006)
IT-Intensive Sectors: Number of Establishments			0.048***	0.056***				0.057***
			(0.010)	(0.012)				(0.013)
Log (Number of IT Patents)					0.003		0.003	-0.000
					(0.004)		(0.004)	(0.003)
Kauffman High-Growth Entrepreneurship						0.001	0.001	0.000
						(0.005)	(0.005)	(0.004)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	5595	5595	5595	5595	5595	5595	5595	5595
R^2	0.746	0.746	0.751	0.752	0.746	0.746	0.746	0.752
Mean Outcome	10.937	10.937	10.937	10.937	10.937	10.937	10.937	10.937

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on log mean wage, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2000 to 2004. All MSA-level regressors are measured at a one-year lag relative to the outcome variable. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table A.4: Explanatory Local Variables and IT Wage Spreads (Excluding Tech Hubs), 2013 – 2019

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	-0.062	-0.064	-0.053	-0.055	-0.062	-0.077	-0.077	-0.070
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.058)	(0.058)	(0.059)
Log (Total Population)	0.024***	0.027***	-0.055*	-0.056*	0.021***	0.018***	0.017***	-0.072**
	(0.005)	(0.006)	(0.029)	(0.029)	(0.006)	(0.004)	(0.005)	(0.030)
Share of College-Educated Adults	-0.068	-0.068	-0.177	-0.183	-0.093	0.043	0.032	-0.088
	(0.101)	(0.101)	(0.117)	(0.119)	(0.104)	(0.093)	(0.101)	(0.118)
IT-Intensive Sectors: Employment HHI		0.004		0.005				0.004
		(0.004)		(0.004)				(0.004)
IT-Intensive Sectors: Number of Establishments			0.075***	0.080***				0.087***
			(0.027)	(0.027)				(0.029)
Log (Number of IT Patents)					0.002		0.001	0.001
					(0.003)		(0.002)	(0.002)
Log (Number of New Businesses)						0.020***	0.020***	0.021***
						(0.005)	(0.005)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10848	10848	10848	10848	10848	10848	10848	10848
R^2	0.122	0.123	0.125	0.126	0.123	0.131	0.131	0.135
Mean Outcome	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.720

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2013 to 2019. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table A.5: Explanatory Local Variables and IT Wage Spreads (Excluding Tech Hubs), 2005 – 2012

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	-0.064	-0.068	-0.064	-0.067	-0.064	-0.070	-0.070	-0.072
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.053)	(0.053)	(0.052)
Log (Total Population)	0.020***	0.025***	0.019	0.017	0.021***	0.019***	0.020***	0.013
	(0.004)	(0.006)	(0.029)	(0.029)	(0.005)	(0.004)	(0.005)	(0.028)
Share of College-Educated Adults	0.125	0.113	0.122	0.097	0.129	0.157*	0.169	0.138
	(0.094)	(0.094)	(0.111)	(0.114)	(0.099)	(0.093)	(0.106)	(0.109)
IT-Intensive Sectors: Employment HHI		0.006		0.007				0.007
		(0.005)		(0.005)				(0.005)
IT-Intensive Sectors: Number of Establishments			0.002	0.008				0.012
			(0.027)	(0.027)				(0.025)
Log (Number of IT Patents)					-0.000		-0.001	-0.001
					(0.002)		(0.003)	(0.002)
Log (Number of New Businesses)						0.007*	0.007*	0.007*
						(0.004)	(0.004)	(0.004)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	10355	10355	10355	10355	10355	10355	10355	10355
R^2	0.143	0.144	0.143	0.145	0.144	0.145	0.145	0.146
Mean Outcome	0.698	0.698	0.698	0.698	0.698	0.698	0.698	0.698

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2005 to 2012. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).

Table A.6: Explanatory Local Variables and IT Wage Spreads (Excluding Tech Hubs), 2000 – 2004

Dependent Variable	**Log (Annual Wage in USD)**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Annual Wage in USD)	0.153***	0.155***	0.155***	0.153***	0.154***	0.153***	0.154***	0.152**
	(0.039)	(0.039)	(0.040)	(0.040)	(0.039)	(0.056)	(0.056)	(0.058)
Log (Total Population)	0.011**	0.019***	0.015*	0.015*	0.014**	0.011*	0.014**	0.015
	(0.005)	(0.006)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.011)
Share of College-Educated Adults	-0.053	-0.028	-0.042	-0.037	-0.026	-0.035	-0.004	0.011
	(0.125)	(0.123)	(0.124)	(0.123)	(0.127)	(0.135)	(0.128)	(0.132)
IT-Intensive Sectors: Employment HHI		0.010**		0.011**				0.012*
		(0.005)		(0.005)				(0.007)
IT-Intensive Sectors: Number of Establishments			-0.005	0.005				0.009
			(0.008)	(0.009)				(0.010)
Log (Number of IT Patents)					-0.002		-0.002	-0.003
					(0.002)		(0.003)	(0.003)
Kauffman High-Growth Entrepreneurship						-0.004	-0.004	-0.005
						(0.005)	(0.005)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓	✓
No. Obs.	5595	5595	5595	5595	5595	5595	5595	5595
R^2	0.205	0.206	0.205	0.207	0.205	0.205	0.206	0.208
Mean Outcome	0.695	0.695	0.695	0.695	0.695	0.695	0.695	0.695

Notes: This table presents coefficient estimates from equation 2 which estimates the effects of MSA-level determinants on the 75th-to-10th-percentile wage spread, among IT occupations within 137 large MSAs excluding the tech hubs, in the time period from 2000 to 2004. All MSA-level regressors are measured at a one-year lag relative to the outcome variable, except for the log mean wage which is measured in teh same year. All columns show coefficients on two agglomeration variables – log total population, and share of college-educated adults. Columns 2–4 also show coefficient estimates on IT-intensity variables, and columns 5–7 also show coefficient estimates on innovation and entrepreneurship variables. Columns 8 show coefficient estimates on all explanatory factors within the same regression. All regressions control for occupation-year fixed effects, as well as MSA-level demographic profiles and economic conditions – log median income, unemployment rate, share of population below poverty line, share of population in each of the age buckets: 20–34, 35–44, 45–54, 55-64, and 65+, as well as share of population in each of the non-mixed ethnic groups: white, black, native American, and Asian. We calculate standard errors using two-way clustering, at the occupation-year and MSA levels (except for columns 6–8 with geographic clustering at the state-level instead, since entrepreneurship variables are measured by state).