SKILL-BIASED TECHNICAL CHANGE AND REGIONAL CONVERGENCE*

Elisa Giannone†

March 2021

Abstract

Between 1940 and 1980, the wage gap between poorer and richer US cities shrank at an annual rate of 1.4% but declined to 0% after 1980. I find that this change was driven solely by highly skilled workers. Based on this evidence, I build a dynamic spatial equilibrium model to quantify the joint contribution of technology and local agglomeration to wage convergence and its decline. Two main results arise: i) a national skill-biased technical change shock explains 50% of the decline in regional convergence; ii) at national level, if technology was not skill-biased, there would be less aggregate inequality but also less growth.

*I am indebted to Erik Hurst, Ralph Ossa, Nancy Stokey, and Jonathan Dingel for their invaluable guidance and support throughout this project. I also want to thank Big Banternghansa, Gady Barlevi, Robert Barsky, Paco Buera, Lorenzo Caliendo, Jeff Campbell, Rebecca Diamond, Cristian Dagnino, David Dorn, Jason Faberman, Pablo Fajgelbaum, Cecilie Gaubert, Alejandro Justiniano, Munseob Lee, Kerstin Holzheu, Greg Kaplan, David Nagy, Nuno Paixao, Steve Redding, Esteban Rossi-Hansberg, and many other seminar participants. I thank those at the University of Chicago, LSE Geography, Marco Fanno first Meeting, the Econometric Society World Congress in Philadelphia, Chicago FED, Junior Scholar Conference at the Minneapolis FED, UIUC, FSU, St. Louis FED, Toronto Rotman, Princeton, Georgetown, UVA-Darden, NYU, Penn State, IIES, Rochester, USC-Marshall, USC, UCLA, SOLE, NBER Macro across Space and Time, SED-Edinburgh, Queen's, Maryland, CEPR-EBRD and Barcelona GSE Summer Forum, NBER Urban, BU, Minneapolis FED and Berkeley. I acknowledge hospitality from the Chicago Fed where part of this work was completed. A special thanks to Che-Lin Su who taught me a lot about numerical methods. Any errors are my own.

†Pennsylvania State University. Email at: elisa.giannone@gmail.com

1 Introduction

Technological innovation was a major driver of changes in the labor market at the national level in the twentieth century. At the local level, for the majority of the twentieth century, poorer US cities had grown faster and were catching up with richer cities by consistently shifting technology from the rich to the poor. However, in the 1980s, simultaneously with the technological innovation shock that favored highly skilled workers, cities with higher share of such workers like Boston, San Francisco, and New York began to pull away from their poorer counterparts. The clustering of highly skilled workers amplified the geographic differences. This growing gap, in turn, gave them less incentive to move away from such cities which reduced the overall mobility rate. Despite an extensive literature has analyzed the impact of technological innovation at the national level, curiously, little has studied its impact at the regional level and how this force, operating through cities, has sequentially affected aggregate macroeconomic patterns of growth and inequality.

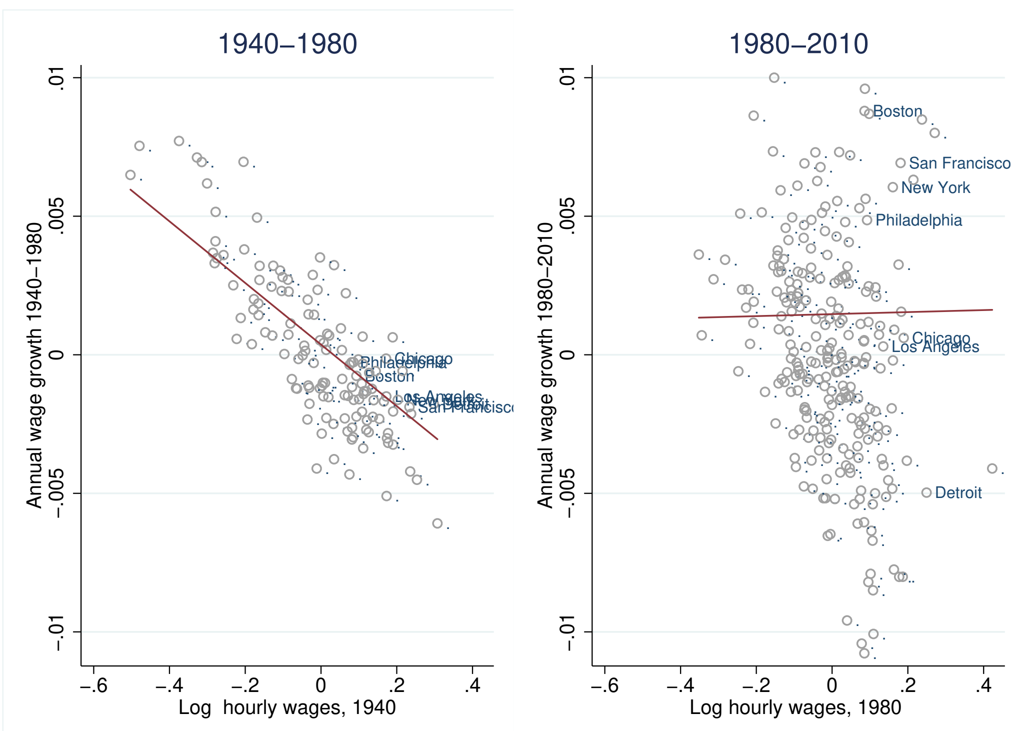
Against this background, two main questions arise. What is the role of technology evolution in explaining regional convergence and its end? And what is the impact of a skill-biased technical change shock on aggregate inequality and economic growth? The main contributions of this paper are to show that technology rationalizes salient changes in the regional data of the last 70 years as well as that it establishes a trade-off between aggregate growth and inequality. The main empirical finding that motivates the rest of the work is that regional convergence only stopped for highly skilled workers but continued for less skilled ones. On the quantitative side, the main findings are that: (i) at the regional level, a national skill-biased technological change (SBTC) interacting with local agglomeration forces explains approximately 50% of the overall decline in convergence; (ii) at the aggregate level, SBTC simultaneously contributes to increasing economic inequality but also growth because of the increase in sorting across space.[1] According to the model, technology diffusion leads to wage convergence across cities. At the same time, a national SBTC shock together with stronger agglomeration economies for the skilled act as a force of divergence. As SBTC unfolds over time, the divergence force starts dominating the convergence force, rationalizing the key fact documented in the data. The model also delivers that the decline in convergence is associated with increasing inequality and higher aggregate growth.

SBTC is the term commonly used in the literature to identify technological innovation that

[1]Other related findings suggest that technology rationalizes both the "Great Convergence" and "Great Divergence" of skills before and after 1980 and the secular decline in geographic mobility. Following the definition of "Great Divergence" coined by Moretti (2012), I define "Great Convergence" of skills, the period between 1940 and 1980, where the relationship between initial college ratio and college ratio growth was negative.

increased the relative demand for highly skilled workers after 1980 (Katz and Murphy 1992). In turn, SBTC has led to a growth in earnings inequality (e.g., Card and DiNardo 2002, Levy and Murnane 1992, Bound and Johnson 1992). National skill premia show similarly timed patterns as regional convergence. In fact, at the regional level, between 1940 and 1980, wages grew 1.4% faster per year faster in poorer US cities rather than in richer cities.[2] This wage convergence ended in 1980, and from then to 2010 wages grew at similar rates in cities with different income levels. Figure 1 plots the annual demeaned average wage growth against its initial demeaned wage level in logs. The β-convergence rate, given by the slope of the trend line, is 0.014 between 1940 and 1980 but goes to zero and is not statistically significant between 1980 and 2010.[3]

Figure 1: Wage Convergence across Cities before and after 1980



Note: This figure plots each city's (demeaned) annual average wage growth against its (demeaned) initial wage level. The left side depicts 1940-1980; the right side depicts 1980-2010. The red line depicts a weighted least square bi-variate regression. Data come from the decennial US Census and the 2010 American Community Survey.

The paper consists of three main parts. In the first part, I identify a *key* fact regarding regional

[2]Throughout the paper, I use "cities" to refer to "Metropolitan Statistical Areas", the geographical unit used in this paper. A formal definition is provided in section 2.

[3]Berry and Glaeser (2005) are the first to point to this decline in convergence across cities after 1980. Ganong and Shoag (2017) show a similar decline in convergence for income per capita across US states after 1980.

convergence followed by two facts about the evolution of skill premium and migration patterns across U.S. cities during the last 70 years. The *key* empirical contribution of this paper is to show that the end of wage convergence occurred only for highly skilled workers. As shown in figure 2, prior to 1980, the wage convergence rate is the same for both highly and less skilled (less-than college-educated) workers. Since 1980, the wages of less skilled workers have continued to converge at 1.4% annually, while the wage convergence rate for highly skilled workers has been close to 0%. Thus, any account of the end of convergence must distinguish between skill groups. The second fact is that after 1980, the correlation between the college ratio and the skill premium is positive across cities.[4] In other words, the relative price of skills has become positively correlated with its relative quantities across cities. Specifically, the third fact is that since 1980, migration destinations of highly skilled workers have shifted towards already skill-concentrated cities, opposite to what previously observed. These facts indicate that performance differences between highly skilled and less skilled workers played a crucial role in the cessation of the regional wage convergence. Due to differences in their initial skill composition, cities might have been more impacted by the national SBTC. These facts are also consistent with the idea that demand grew relatively more than supply in cities with bigger initial concentration of highly skilled workers, benefiting of higher wage growth.

Motivated by this evidence, in the second part of the paper, I develop a theoretical framework that rationalizes both regional convergence and its end for the highly skilled workers. The model features as main ingredients spatial technology diffusion, national skill-biased technical change, local agglomeration spillovers and costly migration. To get quantitatively closer to the data and compare with other stories, multi-industries and housing are included as well. Cities differ along local amenities and an initial component of productivity. Specifically, labor productivity varies across these local labor markets and it is determined by the endogenous agglomeration effects that depend on city-level population and skill concentration, a national skill-biased technical change component and the city-specific technology diffusion process. The model combines two streams in the literature. First, I follow the literature pioneered by Desmet and Rossi-Hansberg (2014) and Desmet et al. (2018) and I introduce spatial technology diffusion. Second, in the spirit of Rosen (1979), Roback (1982), and Diamond (2016), I endogenize household location choice across heterogeneous cities.

[4] I define the college ratio (or skill ratio) as the share of college-educated workers to less-than college-educated workers. The skill premium is defined as the difference between the wage of the college-educated workers and the wages of the less-than college-educated workers.

Accounting for the regional trends and shifts in wages, employment and migration patterns for each skill group requires a combination of the two groups of models above. As I show in Appendix E, spatial equilibrium models with no spatial knowledge diffusion produce divergence rates across US cities but not the convergence rates in wages and skills. They also produce a secular increase in migration rates. Simultaneously, adding skill heteroneity in a growth model such as Desmet and Rossi-Hansberg (2014) allows to explain the differential sorting patterns and the shift we observe in the 1980s. Overall, allowing for heterogeneity in skills and spatial diffusion of productivity over time produces a suitable framework to analyze the core question of this paper; that is, to explain the sharp regional changes that occurred in the long-run and how they mapped to aggregate patterns in growth and inequality. Convergence forces enter through the technology diffusion process as in Desmet and Rossi-Hansberg (2014) while the key divergence forces are SBTC and local agglomeration.[5]

The mechanism works as follows. Spatial technology diffusion pushes for convergence equalizing the wages across cities. Instead, national technological innovation that interacting with local agglomeration forces counterbalances the convergence that existed before 1980. If technology was not skill-biased, the convergence forces would favor the poorer cities by pushing them toward the productivity frontier. The interaction between national SBTC and local agglomeration economies leads to a larger skill premium in more educated locations. Highly and less skilled workers have some degree of complementarity, thus, agglomeration effects raise the wages of all workers, although at different speed. This differential increase in the wages of highly skilled workers changes the relative migration returns for these two groups. Therefore, highly skilled workers migrate to highly skilled cities more than less skilled workers do. Migration has a twofold effect. On the one hand, as more workers migrate to a location, the marginal productivity of each type decreases which decreases the incentive to migrate. On the other hand, when more highly skilled workers move to a location, productivity goes up because of the agglomeration effects, which raises the wages of all the workers, especially those that are highly skilled.

In the third part of the paper, I apply the model to the data. The main goal is to estimate the parameters that govern the productivity process. A key element of the process is a national measure of SBTC. To do so, I average out, at national level, the local exposure to routinization built by Au-

[5] To match the data and to account for other potential factors, I also introduce other divergence forces such as costly migration and housing. Ganong and Shoag (2017) propose a mechanism based on housing regulations to explain the decline in wage convergence. I also compute how much convergence there would exist in the model if I did not account for housing.

tor and Dorn (2013).[6] To estimate the key parameters of the model, especially the remaining ones in the productivity process, I use a GMM procedure. I use equilibrium conditions as moment conditions. To identify such parameters I use as instruments the local exposure to routinization as in Autor and Dorn (2013) as well housing restrictions.[7] Building on existing models that use a related estimation strategy, I extend the estimation strategy to take into account the growth component of the model and the spatial technology diffusion. The identifying assumption is that these proxies for SBTC shocks interact with land unavailability and housing regulations are orthogonal to changes in local productivity. I then calibrate the model using estimated parameters, supplemented with others borrowed from the literature, and solve it numerically.

I validate the model by comparing the patterns of nominal and real wage convergence in the last 70 years in the data and the model itself when considering a national measure of skill-biased technical change. The model matches very closely the patterns of wage convergence and its end for highly skilled workers after 1980 and the continuation for less skilled workers. Next, I construct counterfactual exercises by "turning off", step-wise, the divergence forces in the model. The results show that approximately 50% of the observed decline in wage convergence among highly skilled workers is due to technology becoming more skill-biased. Further, the decomposition of this exercise for highly skilled and less skilled workers indicates that if no shock had occurred, convergence would have been, on average, higher for highly skilled workers than for less skilled workers. I continue validating the model by checking other moments of the data. At the local level, besides matching the features of wages, the model matches the "Great Convergence" and "Great Divergence" in the skill ratio as well as the trends in the secular decline in migration rates. On a related note, the model also matches the increase in wage dispersion across cities, as documented by Hsieh and Moretti (2015), and shows that SBTC explains the biggest share of the increase in spatial wage dispersion.

Finally, at the national level, the model suggests that if a national SBTC shock had not operated through this sorting channel, real wage growth and skill premium would have been smaller than

[6] Autor and Dorn (2013) study the effect of routinization on the polarization of employment and wages. They also argue that their shock fits the overall increase in the skill premium.

[7] I exploit the differential exposure of cities to computerization as in Autor and Dorn (2013). They, in fact, analyze the effect of computer innovation on the output differences across regional labor markets. The national arrival of computers mostly affects occupations that are very routinized because machines can replace those workers. Therefore, the effect of computers is heterogeneous across locations depending on the share of highly and less skilled workers holding very routinized occupations. To capture the exogenous component in the productivity changes of the workers that does not depend on contemporaneous occupational structure, I use 10-year lagged city's composition of routine intensive occupations by industry.

they are. This is an important result that might have normative implications since it assesses how space impacts a trade-off between inequality and growth.

This paper speaks to three strands of the literature. The most related works are recent studies that use dynamic quantitative spatial equilibrium models such as Desmet and Rossi-Hansberg (2014), Nagy (2016), Desmet et al. (2018), Caliendo et al. (2019) and Lyon and Waugh (2019). My paper contributes to this literature by adding both divergence and convergence forces in a unified framework with rich features such as heterogeneous skills, industries, housing, and rich agglomeration forces. Thus, it provides a benchmark to perform regional and aggregate long-run growth and inequality analysis within and across countries. This framework can be used to address questions on the local and aggregate impact of sorting across locations within and across countries. At the same time, it is tractable enough to be estimated.

Furthermore, this paper is related to research that studies the increase in the US spatial dispersion and the "Great Divergence" of skills, such as Berry and Glaeser (2005), Moretti (2012), Hsieh et al. (2013), Eeckhout et al. (2014), Hsieh and Moretti (2015), Diamond (2016), Ganong and Shoag (2017), Baum-Snow et al. (Forthcoming), Fajgelbaum and Gaubert (2018) and Eckert et al. (2020). This paper complements this literature in several ways. Empirically, by finding that regional convergence only stopped for highly skilled workers. Quantitatively, by estimating the importance of a national skill-biased technology shock interacted with local agglomeration in reducing regional convergence, in explaining the "Great Divergence" of skills and other secular changes. Moreover, in a unified framework this paper rationalizes not just the divergence process but also the convergence we used to observe. Finally, this paper contributes to the literature above highlighting the aggregate implications of spatial sorting on the co-movement between growth and inequality through geography.

Additionally, this paper complements the literature on regional convergence across countries and states inspired by the seminal works of Baumol (1986), Barro and Sala-i Martin (1992), and Barro and Sala-i Martin (1997); and continued by Bernard and Jones (1996), Caselli and Coleman (2001), Gennaioli et al. (2014) and Comin and Ferrer (Forthcoming). My paper complements this literature in several dimensions. First, by providing a quantitative model with realistic geography, that can be mapped 1-to-1 to the data. Second, I propose a framework that has both convergence and divergence built-in that can match the data both on prices and quantities and connects the long-standing convergence literature to the one about cities.

The remainder of the paper is organized as follows. Section 2 describes the data and the empirical analysis. Section 3 proposes a theoretical framework. In Section 4, I estimate the core parameters and calibrate the model. In Section 5, I solve the model, show how it maps to the data and conduct a counterfactual analysis. Section 6 concludes with a brief summary and future directions.

2 Data and Empirical Regularities

The main new empirical finding of this paper is that regional convergence stopped only for highly skilled workers but did not for less skilled workers. This fact motivates the exploration of the skill premium and migration patterns for the last 70 years. The facts jointly point to the direction that supply forces (such as marginal return to labor) have been over time dominated by demand forces (such as skill-biased agglomeration and technological change) pushing the skill premium to go up more in already skill abundant cities and more skill migrants to keep joining those same cities. This rationalizes the overall decline in wage convergence. To the best of my knowledge, facts in sections 2.2, 2.3 and 2.4 are new to the literature while fact 2.5 reproduces the previously documented "Great Divergence" and introduces the "Great Convergence".

2.1 Data

My analysis draws on the Census Integrated Public Use Micro Samples (IPUMS) for the years 1940, 1950, 1960, 1970, 1980, 1990, and 2000; and the American Community Survey (ACS) for 2010 (Ruggles et al. (2015)).[8] In order to construct measures of migration, I use the March Current Population Survey (CPS) data that is a monthly US household survey conducted jointly with the US Census Bureau and the Bureau of Labor Statistics. The focus is on household and demographic questions. I use measures of geographic constraints and land use regulations from Saiz (2010). More details about the data and the definitions of the variables are in the appendix.

2.2 The End of Wage Convergence for Highly Skilled Workers

The main empirical fact of this paper shows that regional wage convergence stopped only for highly skilled workers. Figure 2 shows that the cross-MSA wage convergence rates between 1940 and 1980 were the same for highly skilled and less skilled workers. But, they differ strongly after 1980. Between 1980 and 2010, the wage convergence rate occurred only among less skilled workers but not for highly skilled workers.

[8]The Census samples for 1980, 1990, and 2000 include 5% of the US population; 1970 Census and ACS sample includes 1% of the population; and 1950 Census sample includes approximately 0.2% of the population.

To illustrate these patterns, I run the same "convergence" regression as in Baumol (1986):

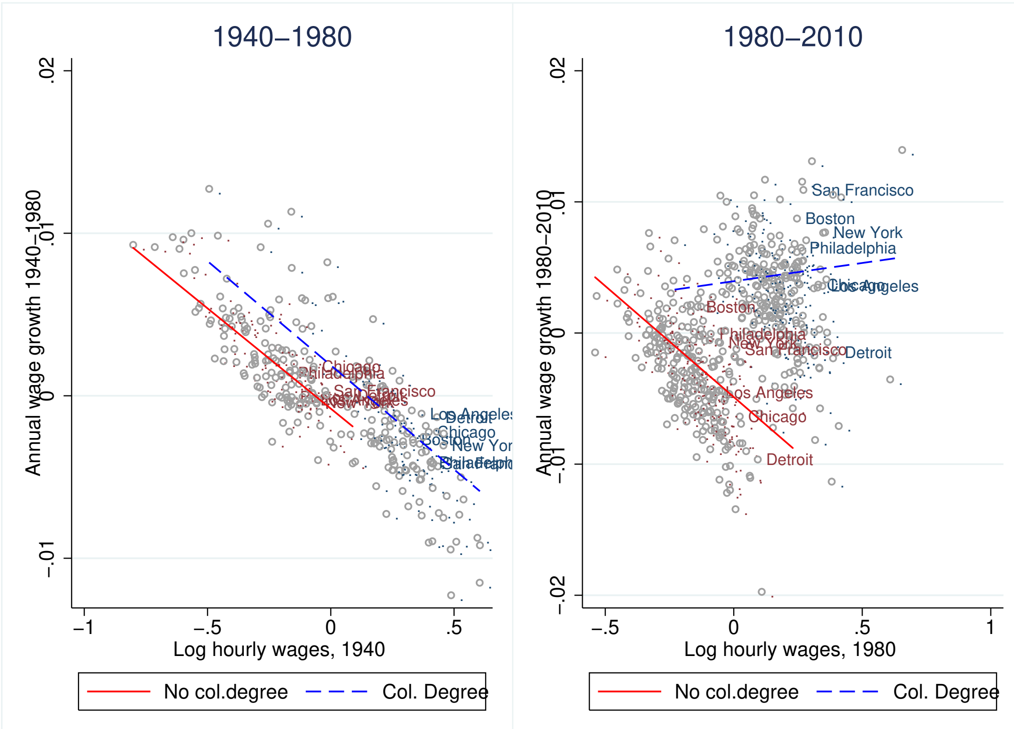
$$\frac{w_{kjt} - w_{kj\tau}}{(t - \tau)} = \alpha + \beta^k w_{kj\tau} + \epsilon \tag{1}$$

where k is the skill group, highly skilled H or less skilled L; j is the MSA; and t is the final year of the analysis and τ is the initial year. $w_{kj\tau}$ is the log hourly wage by skill group k in MSA j at time τ. The dependent variable is the annual average wage growth of log hourly wages between τ and t. All the regressions are weighted by the initial population size. If the estimates of β^k are negative and statistically significant, then there is wage convergence and the convergence rate is exactly β^k. If they are positive and statistically significant, then there is wage divergence. In Figure 2, I plot the observations at the MSA level by skill group k and then the line fit, where β^H and β^L-convergence rates are the slopes of the lines. The blue dashed line is the β-convergence for L, and the red solid line is the β-convergence for H. Each circle is an observation by MSA and skill group. I label the 10 biggest US MSAs in red for the observation of the less skilled and in blue for highly skilled workers, respectively.

Between 1940 and 1980, there was no difference between cross-MSAs wage convergence rates, β^H and β^L. Between 1980 and 2010, the convergence rate β^L was still negative and statistically significant, but β^H was not. This difference means that the end of convergence was driven only by the wages of highly skilled workers, because the wages of less skilled workers still converged across MSAs. In Panel B of Table 1, I report the estimates of β^L and β^H in the two different periods both for population-weighted and non-population-weighted regressions. For the population-weighted regression, β^L and β^H are, respectively, -0.0123 and -0.0143 between 1940 and 1980. Both estimates are statistically significant. However, the estimates of β^L and β^H between 1980 and 2010 are respectively, -0.0169 and 0.000636. The estimate of β^L is statistically significant but the estimate of β^H is not statistically different from zero. In the appendix, I run several robustness tests for this fact. First, I estimate the rolling convergence for the highly skilled and less skilled workers separately for 10- and 20-year windows. Second, I run the same regression as above for compositionally adjusted wages.[9]

[9]The results are very robust to different specifications.

Figure 2: Wage Convergence across MSAs before and after 1980 by Skill Group



Note: This figure plots each MSA's annual average wage growth (demeaned) against its (demeaned) initial wage level by skill type (highly skilled and less skilled workers). The left depicts 1940-1980; the right depicts 1980-2010. Each MSA's circle size is proportionate to its initial population size by skill group. The red solid and the blue dashed line in each graph depict a weighted least square bi-variate regression, respectively, for less and highly skilled workers.

2.3 Skill Premium By Skill Ratio Over Time

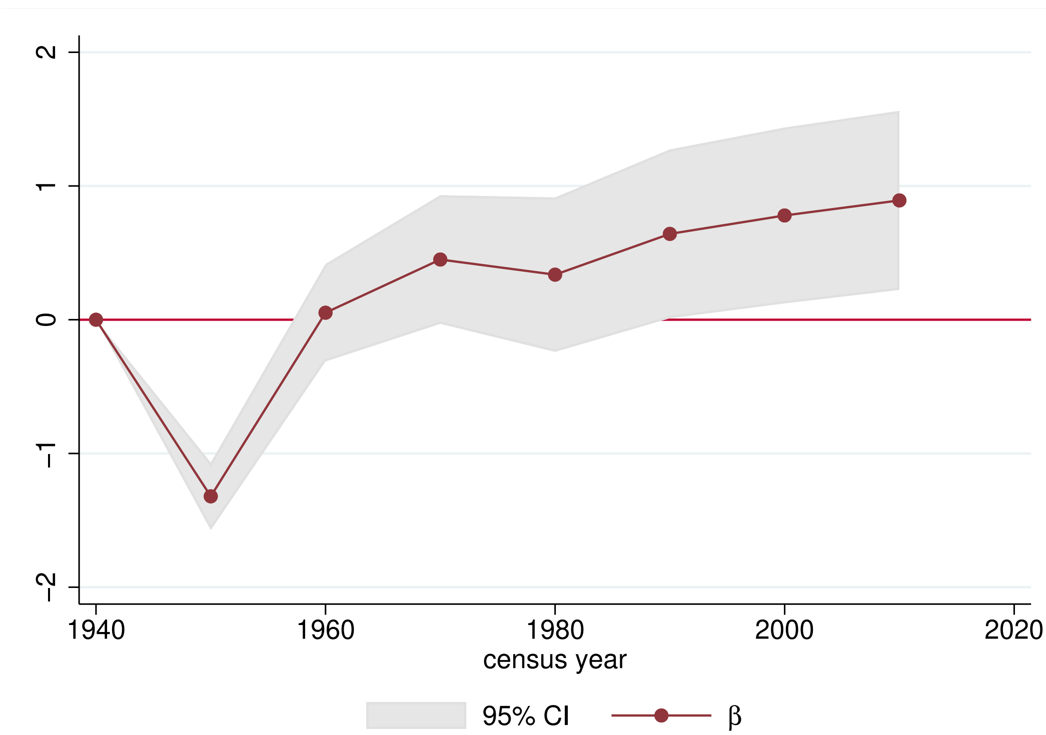
The second empirical fact shows that while skill premium used to be lower in skill abundant places, in recent years it has become higher in skill abundant places. In figure 3 I define skill premium as the difference between the wages of the workers that are highly skilled and the workers that are less skilled. I run the following regression:

$$\ln\left(\frac{\hat{w}_{Hjt}}{\hat{w}_{Ljt}}\right) = \sum_{t=1940}^{2010} \beta_t \left(\frac{H_{jt}}{L_{jt}}\right) + \phi_j + \phi_t + \epsilon_{jt} \tag{2}$$

where $\hat{w}_{Hjt}$ and $\hat{w}_{Ljt}$ are the compositionally adjusted wages for MSA j at time t respectively for highly skilled and less skilled workers. ϕ_j is the MSA fixed effect, and ϕ_t is the time fixed effect. $\frac{H_{jt}}{L_{jt}}$ is the ratio of the total number of workers that are highly skilled to that of less skilled workers in MSA j at time t. I run the regression for $t = \{1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010\}$.

Once I run the regression for each year of the Census, I plot the estimate for the coefficient β_t for each year. This coefficient can be interpreted as an increase of one standard deviation in $\frac{H_{jt}}{L_{jt}}$ that is going to affect the skill premium by β_t standard deviations. In figure 3, there is a clear pattern for the growth of the skill premium by MSA education. In Table 2, I report the estimates of β_t that control for population. Baum-Snow and Pavan (2013) find that at least 23% of the overall increase in the variance of log hourly wages in the US from 1979 to 2007 is explained by the more rapid growth in the variance of log wages in larger locations relative to smaller locations after controlling for the skill composition of the workforce across MSAs of different sizes. This evidence reinforces the presence of growing agglomeration economies and motivates the decision to introduce them in the theoretical framework, both for population and for skill-ratio.

Figure 3: Skill Premium by MSA Education Levels



Note: This figure plots the estimate of the coefficient β for the regression 2. On the horizontal axis, I have the decades from 1940 to 2010. While, on the vertical axis, I have estimates of coefficient β for each decade from 1940 to 2010. Moreover, there is a line starting at zero on the vertical axis.

2.4 The Concentration of Highly Skilled Migrants

The third fact suggests that a higher share of highly skilled workers is moving more and more to highly skilled places. The research on migration has proven that educated workers migrate more than less-educated workers. But where are they actually migrating to? Are they migrating to less educated places to take advantage of the scarcity of a highly skilled labor force? In order to assess which type of workers migrate more to highly educated MSAs, I run a difference-in-difference analysis as in equation (3).

$$1\left(\text{Migrant}_{ijt}\right) = \alpha + \beta 1\left(H_{ijt}\right) + \gamma \frac{H_{jt}}{L_{jt}} + \sum_{t=1963}^{2013} \delta_t 1\left(\mathrm{H}_{ijt}\right) * \left(\frac{H_{jt}}{L_{jt}}\right) + \Gamma X_{ijt} + \phi_j + \phi_t + \mu_{ijt} \quad (3)$$

The dependent variable in this equation is whether worker i in MSA j at time t is a migrant or not. The variable equals one if the worker is a migrant. On the right-hand side, there is an indicator variable H_{ijt} that equals one if the worker is highly skilled and zero otherwise. The second variable is the skill ratio $\frac{H}{L}$ in each MSA at each time. Third, there is the interaction between the first two variables. Regression 3 also includes MSA and time fixed effects. I use the estimated coefficient δ_t to compute the marginal effect of being a highly skilled worker and being in a more skilled MSA on the probability of being a migrant. The X_{ijt} represents the economic demographics of the workers such as age, gender, race, and nationality.[10]

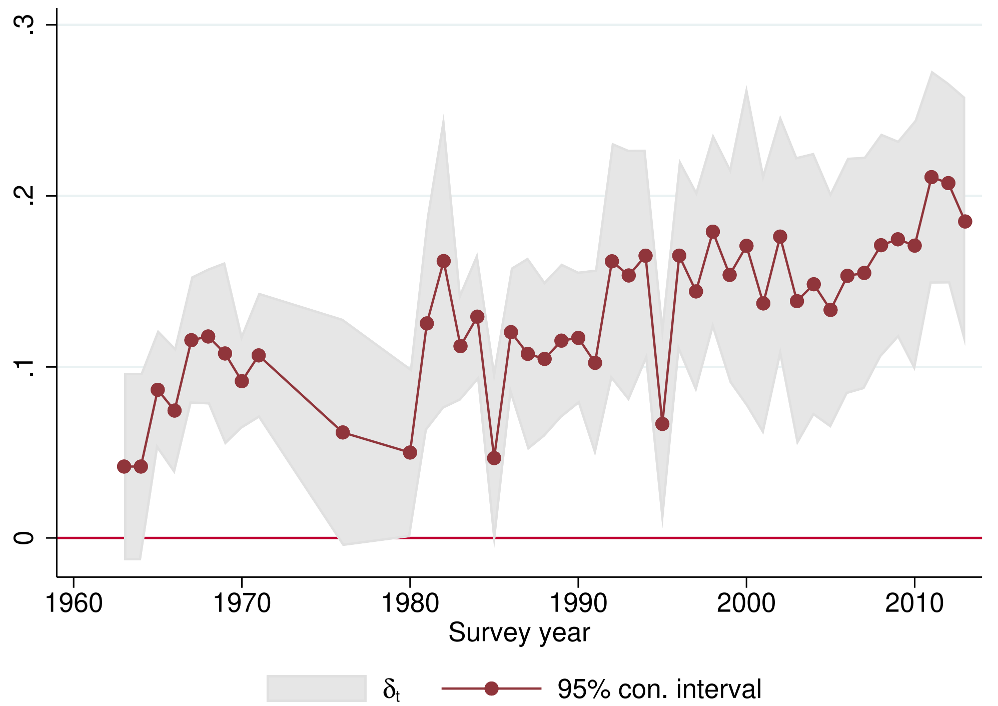
I run regression 3 both as a linear and a logit model. I focus on the marginal effect of δ_t to find the impact of the probability of worker i in MSA j at time t of being a migrant or not given MSA j's skill ratio that is interacted with the worker being highly skilled. I run the same regression with the CPS where the information about the migration status of the worker is available for the years from 1962-2010 except for 1972-1975 and 1976-1979. In the appendix, I run the same exercise using Census data extracted from the IPUMS. Each observation in figure 4 corresponds to the coefficient δ_t in regression (3). I use this as a robustness check. Then, to make evaluations consistent with the Census data and to rule out potential biases because of the cycles, I take the average of the estimate for each decade for the available data. For instance, for the 1960s, I take the average of the data available up to 1965. For the 1970s, I take the average of the estimates from 1966 to 1975 and so on.

Figure 4 shows that the marginal propensity to migrate conditional on being a highly skilled worker and moving to a highly skilled MSA increases over time. Thus, in relative terms highly skilled workers concentrate more and more over time in the more educated MSAs. This finding goes well in accordance with the hypothesis that highly skilled workers concentrate more and more

[10]The more detailed description is the same as the one I did for the compositionally adjusted wages.

in educated MSAs. Table 3 shows the evolution over time of the marginal effect of being highly skilled and being in a highly educated MSA on being a migrant.

Figure 4: Migration Rate by Destination Education Level



Note: This figure plots the estimate of the coefficient δ for the regression 3. On the horizontal axis, I have the years from 1962 to 2010. While, on the vertical axis, I have estimates of coefficient δ for each year from 1962 to 2010. Moreover, there is a line starting at zero on the vertical axis.

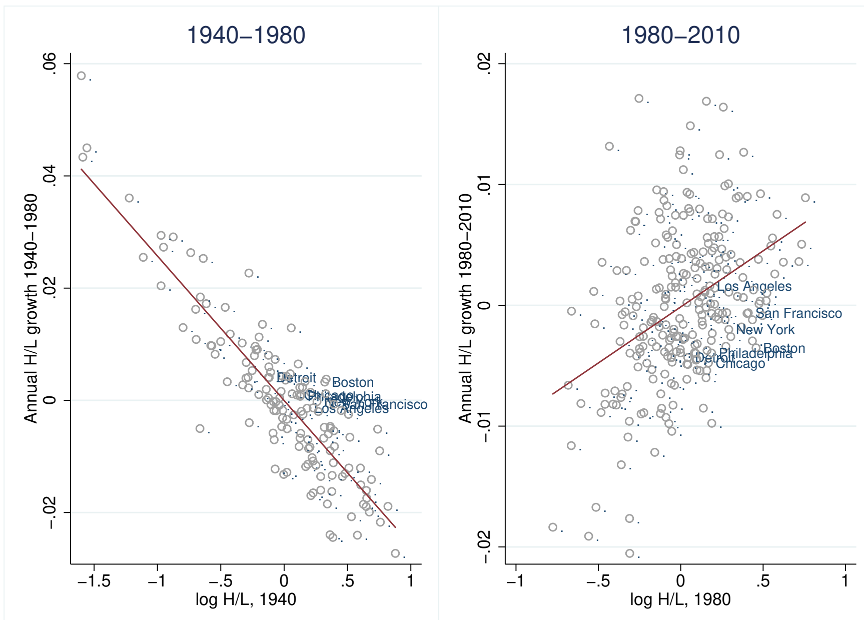
2.5 Skill Ratio: The Convergence and "The Great Divergence" after 1980

The last fact of this paper shows that before 1980 there was convergence in skill ratio and it replicates the "Great Divergence" of skills after 1980. Moretti (2004), Berry and Glaeser (2005), Moretti (2012) and Diamond (2016) show that the skill ratio of workers between 1980 and 2010 was diverging across MSAs. Moretti (2012) coins the term "The Great Divergence" to stress how the skills diverge over space. But what happened to the skill ratio before 1980? Was the skill distribution converging across MSAs when wages were converging? To answer this question, I analyze the convergence rates of the skill ratio over the last 70 years. I estimate the following specification:

$$\ln\left[\frac{H_{jt}}{L_{jt}} - \frac{H_{j\tau}}{L_{j\tau}}\right]\frac{1}{(t-\tau)} = \alpha + \beta^{skill}\ln\frac{H_{j\tau}}{L_{j\tau}} + \epsilon \tag{4}$$

where H_{jt} and L_{jt} are, respectively, the number of highly and less skilled workers living in MSA j at time t and the initial period τ. The dependent variable is the average annual growth of the skill ratio between τ and t. With this regression, I can assess the extent to which growth in the skill ratio is related to the initial skill ratio. This regression is analogous to the regressions run in Figures 1 and 2 but for quantities rather than for wages. I run this regression over different periods using the Census and ACS data. In Figure 5, I plot the observations at the MSA level and then the line fit, where β^{skill}-convergence rates are the slope of the lines. Each circle is an observation by MSA. I label the 10 biggest US MSAs. Between 1940 and 1980, the β^{skill}-convergence rate was negative and statistically significant, therefore, the term "Great Convergence". However, as observed in Moretti (2004), between 1980 and 2010, the β^{skill}-convergence rate was positive and statistically significant that indicates divergence. Table 4 has the results from decomposing the years in shorter periods. The results show that the distribution of highly skilled and less skilled workers across MSAs was converging between 1940 and 1980 and then started to diverge between 1980 and 2010. Panel A shows the results when the difference between t and τ is 10 years. While in Panel B, the same difference is set at 20 years. Panel A shows that the estimated coefficients are negative and statistically significant until 1970, then they become not significant for 1970-1980 and 1980-1990. Further, between 1990-2000 and 2000-2010 they become positive and statistically significant. A 1% increase in the college share ratio increases the change in the college share by 0.07% and 0.04%, respectively, between 1990-2000 and 2000-2010. In Panel B, the results are quite similar, but in column (1) the coefficient is positive and statistically significant. That coefficient is actually calculated for 1940 to 1970 since data for 1960 is not available. Therefore, in a 30-year time span, the results should have reversed for other reasons. But, the coefficient between 1950 and 1980 is negative and statistically significant as expected. Specifically, a 1% increase in the college ratio in 1950 decreases the change in the college ratio between 1980 and 2010 by 0.32%.

Figure 5: Skill Convergence across MSAs before and after 1980



Note: This figure plots each MSA's annual average skill growth (demeaned) against its (demeaned) initial skill level. The left depicts 1940-1980; the right depicts 1980-2010. Each MSA's circle size is proportionate to its initial population size. The red line depicts a weighted least square bi-variate regression. The line in each graph represents a weighted regression line from the bi-variate regression.

3 A Spatial Equilibrium Model with Heterogeneous Skills and Technology Diffusion

The empirical analysis indicates that long-run changes pushing the concentration of skills and their returns in already highly skilled locations occurred. Specifically, these changes affected patterns of regional convergence, skill premium, and cross-MSA migration. These observations together and the time in which such changes occurred point to the direction that a national skill-biased technology interacting with local agglomeration might rationalize the afore-mentioned long-term changes. But how can these effects be disentangled? How important is each of these mechanisms? How do they affect the aggregate patterns of growth and inequality? To answer these questions, I develop a dynamic model of cities based both on the current spatial equilibrium literature nesting it with spatial economic growth.

I start with a simple two period and two city setup of geography and skills to highlight the mechanism through which a national skill-biased technological change shock interacting with agglomeration pushes divergence over convergence forces. Then, I present a quantitative, general equilibrium model with the objective to rationalize the empirical facts, quantify the role of the main mechanisms, compare with others and, finally, explore the aggregate trade-off between inequality and growth.

3.1 A Simple Two Period Model of Cities and Skills

There are two types k of households, highly skilled H and less skilled L. In each time period t, where $t = \{1, 2\}$ they decide how much to consume and which location j to pick for living. There is a set of $J = \{\text{SF, D}\}$ locations where SF stands for San Francisco and D stands for Detroit. Workers can move between these two locations. The labor of H and L are the only two factors of production. Each worker provides, inelastically, one unit of labor in the location where she lives for which she is compensated with a wage. Workers each periods decide where to live and what to consume. The wage for the H type is derived from the following expression:

$$w_{Hj} = \left(\frac{H_j}{L_j}\right)^{\gamma^H} MRL_{Hj} S_H \xi_j \qquad w_{Lj} = \left(\frac{H_j}{L_j}\right)^{\gamma^L} MRL_{Hj} S_L \xi_j$$

where $\left(\frac{H_j}{L_j}\right)^{\gamma^H}$ are agglomeration forces that depend on the ratio of high-to-low skilled in location j; S_H and S_L are exogenous national-level skill-biased productivities; MRL is the marginal return to labor in the location j, and ξ_j is a productivity that depends from the previous period following a spatial diffusion process such that

$$\xi_{jt} = \xi_{jt-1}^{\gamma^2} \sum_{m \neq j} \omega \xi_{mt-1}^{1-\gamma^2} \tag{5}$$

where ω is the distance between the two locations.

Proposition 1 *A spatial diffusion process as described in 5 pushes for regional convergence.* Suppose that initial productivity at time 0 is higher in San Francisco than Detroit such that $\xi_{SF0} > \xi_{D0}$. Because of the spatial diffusion process in 5, then, $\Delta\xi_{SF-D1} < \Delta\xi_{SF-D0}$. Therefore, $\Delta w_{SF-D1} < \Delta w_{SF-D0}$, where $w = w_H * H + w_L * L$. Therefore, the spatial diffusion process will push for convergence between San Francisco and Detroit.

Now, suppose that there is a national skill-biased technical change shock that increases S_H. The effect on local wage of both San Francisco and Detroit high-skilled workers will be positive since it increases productivities.[11] On top of that, as shown in proposition 2, the same shock might have a larger effect in San Francisco if H is higher there.

Proposition 2 *A national SBTC shock (or a national increase in S_H) will increase regional divergence if agglomeration forces are stronger than convergence forces such as MRL and spatial diffusion.*

But it might differ among locations depending on the share of high-skilled workers. Specifically,

$$\frac{\partial^2 w_{Hj}}{\partial S_H \partial H_j} > 0 \qquad \text{if} \qquad \gamma^H \left(\frac{H_j}{L_j}\right)^{-1} > -\frac{1}{MRL_{Hj}} \frac{\partial MRL_{Hj}}{\partial H}$$

The expression above is key to understand the forces in the model. If agglomeration forces dominate decreasing MRL, then, $\frac{\partial^2 w_{Hj}}{\partial S_H \partial H_j}$ will be positive. Therefore, more high-skill labor H is present in a region, then, a national skill-biased technical shock might increase the wages of high-skilled more in places where they are more concentrated (San Francisco) than where they are less

[11] $\frac{\partial w_{Hj}}{\partial S_H} = \left(\frac{H_j}{L_j}\right)^{\gamma^H} MRL_{Hj} \exp(\xi_j)$. Since all the elements are positive, then, a national skill-biased technical change shock will increase the wage of the high-skills differently in all locations.

concentrated (Detroit). Comparing San Francisco to Detroit where $\left(\frac{H_{SF}}{L_{SF}}\right) > \left(\frac{H_D}{L_D}\right)$, then, if

$$\left(\frac{H_{SF}}{L_{SF}}\right)^{\gamma^H} MRL_{HSF}\text{exp}(\xi_{SF}) > \left(\frac{H_D}{L_D}\right)^{\gamma^H} MRL_{HD}\text{exp}(\xi_D)$$

which depends on the agglomeration forces versus the MRL and ξ, then, H types will be more likely to move to San Francisco than to Detroit exacerbating the differences with Detroit and pushing for divergence. This is the key mechanism that drives regional divergence in the model and for aggregate increase in inequality.

Proposition 3 *A national SBTC shock (or a national increase in S_H) will increase the national skill premium (or income inequality) but will also increase economic growth.*

Proposition 3 highlights the role of agglomeration on the aggregate economy. Since agglomeration forces become larger and larger over time as a result of the interaction with SBTC and sorting, then, economic growth at national level might increase in the long-run.

3.2 Quantitative Model

To quantify the effect of SBTC and local agglomerations and to compare it with other forces, I embed the mechanism above described in a quantitative model of cities over time. This model departs from the one above by adding multiple locations, migration costs, a housing sector, local amenities and industries. This features make the model more quantitative realistic and allow to take into account other potential stories such as housing or sectoral differences. Each location produces a tradable good T, a set of non-tradable intermediates, $d \in D$, and housing O. The production of tradable T employs both highly and less skilled labor. The productivity terms are different for the two sectors' production functions. The endogenous component is a function of the ratio of highly skilled workers to less skilled workers, and population. Moreover, worker productivity is different across locations. The housing supply is a function of the local rents.

3.3 Preferences and agents' choices

In each period, agents derive utility from consuming a tradable good T and housing O according to Stone-Geary preferences with a subsistence level housing $\bar{O}$.[12] Agents also derive utility from

[12]The preferences present a degree of non-homotheticity similar to Ganong and Shoag (2017). This will allow to make a fair quantitative comparison with the housing mechanism. The main qualitative predictions of the model would not change if there was not non-homotheticity.

exogenous amenities A_{kjt} and from living in more highly skilled cities with higher (H_{jt}/L_{jt}) to some exponent γ^p. The period utility of an agent i of type $k \in \{H, L\}$ who resides in location j at time t and lives in a series of locations $\bar{j}_- = (j_0, ..., j_{t-1})$ in all previous periods is given by

$$u_{ikjt\bar{j}_-} = u_{ikjt} \prod_{s=1}^{t} m_k(j_{s-1}, j_s)^{-1}$$

where u_{ikjt} is the utility which depends only on the current location j of the agents; and $m_k(j_{t-1}, j_t)$ is the migration cost of type k from moving from location j_{t-1} to location j_t, which is also a permanent utility loss for moving from j_{s-1} in $s-1$ to j_s in s. The utility u_{ikjt} is given by

$$u_{ikjt} = \theta \ln(T_{kjt}) + (1-\theta)\ln(O_{kjt} - \bar{O}) + A_{kjt} + \gamma^p \ln(H_{jt}/L_{jt}) + \zeta_{ijt}$$

where ζ is a taste shock distributed according to a Gumbel (or Type I Extreme Value) distribution. Thus,

$$\Pr[\zeta_{ijt}] = e^{-e^{(-\zeta_{ijt})}}$$

I assume that ζ_{ijt} is i.i.d. across locations, individuals, and time. Agents discount the future at rate β and so the welfare of an individual i in the first period is given by $\sum_t \beta^t u_{itj\bar{j}_-}$ where j_{it} denotes the location at time t, $\bar{j}_-$ denotes the history of previous locations, and j_{i0} is given. Agents earn a wage W_{kjt} from their work. Every period, after observing their idiosyncratic taste shock, agents decide where to live that is subject to mobility costs m_k. These costs are paid in terms of a permanent percentage decline in utility. I use the same assumption about the separability of moving costs as in Desmet et al. (2018) such that $m_k(s, j) = m_{k1}(s)m_{k2}(j)$ with $m_k(j, j) = 1$ for all $j \in S$. This assumption turns out to be extremely useful for the feasibility of the model because it means that agents' choice of location depends only on current variables and not their location history.[13] Therefore, I rewrite the agents' problem in a recursive formulation. The value function for an agent living in location j after observing a distribution of the taste shock in all locations is

[13] Caliendo et al. (2019) solve the migration problem dynamically by keeping track of the distribution of workers across locations by using a "hat algebra" method. One extension of the current model would be to incorporate that decision on top of the current features. However, in order to use their method, I would need to measure the migration flows across cities in 1940. Unfortunately, these data are not currently available to the best of my knowledge.

given by

$$
\begin{aligned}
V_{kt}(j,\zeta_i') =& \max_{j'}\left[\frac{V_{ikj't}}{m_k(j,j')} + \beta E\left(\frac{V_{kt+1}(j',\zeta_i'')}{m_k(j,j')}\right)\right] \\
=& \frac{1}{m_{k1}(j)}\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')} + \beta E\left(\frac{V_{kt+1}(j',\zeta_i'')}{m_{k2}(j')}\right)\right] \\
=& \frac{1}{m_{k1}(j)}\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')} + \beta E\left(\max_{j''}\left[\frac{V_{ikj''t+2}}{m_{k2}(j'')} + \beta E\left(\frac{V_{kt+2}(j'',\zeta_i'')}{m_{k2}(j'')}\right)\right]\right)\right]
\end{aligned}
\tag{6}
$$

From the last line of equation 6, it follows that the choice of current location is independent of past and future locations. This independence means that the value function can be rewritten, which isolates the current component as a static problem.[14] Thus,

$$
\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')}\right]
$$

After deciding location j', the agent solves the following static consumption problem:

$$
V_{ikj't} = \max_{T_{kj't},O_{kj't}}[\theta \ln(T_{kj't}) + (1-\theta)(\ln(O_{kj't} - \bar{O}_{kj't}) + A_{j't} + \gamma^p \ln(H_{j't}/L_{j't}) + \zeta_{ij't}]
$$

$$
\text{s.t. } T_{kj't} + O_{kj't}R_{j't} = W_{kj't}
$$

The indirect utility of agent i of type k at time t living in MSA j can be written as

$$
V_{ikjt} = \left[\theta \ln(\theta W_{kjt} - R_{jt}\bar{O}) + (1-\theta)\ln\left((1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}\right) + A_{kjt} + \gamma^p \ln\left(H_{jt}/L_{jt}\right) + \zeta_{ijt}\right]
$$

where k is the skill group of the individual, which can be "highly skilled" H_{jt} or "less skilled" L_{jt}. w_{kjt} is the log of the wages for each skill type k in location j at time t.

Using the properties of the Gumbel distribution and following McFadden (1973), I derive the number of workers of types H and L living in each location j at time t.

$$
H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))}
\tag{7}
$$

[14] The derivation above follows from Desmet et al. (2018).

$$L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{8}$$

where

$$\delta_{kjt} = \theta \ln(W_{kjt} - R_{jt}\bar{O}) + (1-\theta)[\ln(1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + A_{kjt} + \gamma^p \ln\left(H_{jt}/L_{jt}\right) \tag{9}$$

3.4 Technology

In the next subsection, I describe the production technology of the final tradable sector, T; the non-tradable intermediates, $d \in D$; and the housing sector, O. The final good is produced using all the intermediates jointly in a CES fashion. The local market produces intermediates and housing. The intermediates are produced using a CES with highly skilled and less skilled labor. The housing sector is produced depending on the price of the housing sector as in Ganong and Shoag (2017). Because the tradable good T is freely tradable across locations, the price of T, $P_{Tjt} = p_{Tjt}$, $\forall j$, that means it is the same across locations and is assumed to be a numéraire.

3.5 Final Good Production

The final good is produced combining all the intermediate d jointly in a CES fashion where the elasticity is given by α, and the share used in the production function is μ_d. Specifically,

$$T_{jt} = \big(\sum_d \mu_d Y_{djt}^{\alpha}\big)^{1/\alpha}$$

3.5.1 Intermediates Sector

The production function in equation 10 is a CES that uses two types of labor H_{djt} and L_{djt} as imperfect substitute inputs.[15]

$$Y_{djt} = [\eta_{Ldjt} L_{djt}^{\rho} + \eta_{Hdjt} H_{djt}^{\rho}]^{\frac{1}{\rho}}, \qquad \forall\, j = \{1, ..., N\} \tag{10}$$

η_{Hdjt} and η_{Ldjt} denote the productivity of H and L, respectively, in sector d at location j for

[15] I do not include physical capital in this model since my focus is on the composition of the labor force and human capital. However, the consequences of including capital might differ depending on whether capital is mobile or immobile.

time t. Productivity is divided into an exogenous and an endogenous component.[16] Departing from the standard formulation of a CES as in Katz and Murphy (1992), I follow the recent literature on agglomeration in order to make productivity dependent on both endogenous and exogenous components. Endogenous differences in productivity depend on the industry mix in the location. As Diamond (2016) argues, the literature on social returns to education has shown that areas with a higher concentration of college graduates are more productive due to knowledge spillover.[17] Adding knowledge spillover through endogenous productivity that derives from the skill ratio is supported also by my empirical findings, as in section 2. These two facts suggest that 1) the higher the skill ratio, the higher the wage premium in the location and 2) highly skilled workers migrate to cities with a higher skill ratio more frequently than do less educated workers. These two facts embrace the hypothesis that knowledge spillover can be higher in cities with a higher concentration of highly skilled workers. Simultaneously, following Davis and Dingel (2014) and Baum-Snow et al. (Forthcoming), the spillover effects also appear with respect to population, not just the skill ratio.[18] To allow for both effects, I write the expressions η_{Hdjt} and η_{Ldjt} as follows:

$$\eta_{Hdjt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (L_{jt}+H_{jt})^{\phi^H} S_{Ht}^{\lambda^H} \exp(\xi_{Hdjt}) \qquad \eta_{Ldjt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (L_{jt}+H_{jt})^{\phi^L} S_{Lt}^{\lambda^L} \exp(\xi_{Ldjt})$$

where S_{kt} is the exogenous skill-biased technology component for $k \in \{H, L\}$.[19] The exogenous productivity component is ξ_{kdjt}. ξ_{kdjt} at time 0 is given and then evolves according to:

$$\xi_{kdjt} = \xi_{kdjt-1}^{\gamma^2} \left[\int_s \omega(j,s)\xi_{kdst-1} ds \right]^{1-\gamma^2} \tag{11}$$

where $\omega(j, s)$ is a symmetric measure of distance between location j and location s and $\gamma^2 \in [0, 1]$.[20] If $\gamma^2 < 1$, then the productivity in location j is dependent on the productivity of the other

[16] Applying a change in variable as in Diamond (2016), Y_{djt} can be rewritten as a function of data (w_{Ljt}, w_{Hjt}, H_{djt}, L_{djt}, H_{jt}, L_{jt}) and parameters (ρ, γ^L, γ^H, ϕ^L, ϕ^H). More details are given in the appendix in section D.3.

[17] In the current version of Diamond (2016), spillovers are not modeled with parametric formulation but more importance is given to utility spillovers. My paper, however, benefits by modeling productivity spillovers with specific functional forms, especially for the counterfactual analysis.

[18] To guarantee the existence of a steady state, I will need to derive sufficient conditions to be imposed on the agglomeration effect.

[19] In the Appendix, I present a version of the model with endogenous SBTC modeled as technology adoption in line with Beaudry et al. (2010). However, this version does not reproduce features that I see in the data, such as correlation between the skill premium and local supply of skilled labor.

[20] As a robustness test, I numerically test this productivity process, holding ω constant such that $\int_S \omega ds = 1$. The

locations. This dependence will introduce convergence into the model through spatial knowledge diffusion.

The profits π of the firm come from the following maximization problem:

$$\pi_{djt} = \max_{l,\,h} p_{djt}[\eta_{Ldjt} l^{\rho} + \eta_{Hdjt} h^{\rho}]^{\frac{1}{\rho}} - W_{Hjt}h - W_{Ljt}l$$

where l and h are, respectively, the amount of less and highly skilled labor used by one firm that produces the intermediate good d. p_{djt} is the price at which the intermediate d is sold. A free entry condition drives profits to zero since the firms keep entering until the profits are equal to zero. Therefore, a firm choosing its production in period t knows that its current and future profits are going to equal zero. This result is extremely useful in solving the model. It means that the dynamic component of the model is equivalent to a repeated static model, which facilitates the numerical solution.

Since the labor markets are perfectly competitive, the wage in each location is equal to the marginal product of labor as shown in equations 12 and 13, which derive the first-order condition of the firms.

$$W_{Hjt} = p_{djt}\eta_{Hdjt}[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}H^{\rho-1}_{djt} \tag{12}$$

$$W_{Ljt} = p_{djt}\eta_{Ldjt}[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}L^{\rho-1}_{djt} \tag{13}$$

3.5.2 Housing Market

The supply of housing is a convex function of its price. The higher the price of housing, the higher the supply.[21]

$$O_{jt} = R^{\mu}_{jt} \tag{14}$$

where the exponent μ represents the elasticity of housing and R is the rental rate of houses in location j at time t. This equation mimics the housing sector following Ganong and Shoag (2017). The idea behind this expression is that regulations affect the elasticity of supply as a direct cost

results are qualitatively unchanged.

[21] To create a fully dynamic housing model with investment decisions along the lines of Glaeser and Gyourko (2006) is a possible extension of the paper. However, to avoid moving the focus of the paper away from skill-biased technology and agglomeration, I keep the housing market as simple as possible. This simplification also enhances comparability to Ganong and Shoag (2017). I run some simulations fluctuating the value of the parameter μ to very large levels and to small levels to check how the housing would respond.

shock. Local housing demand follows from the household problem and is given by:

$$H_{jt}\left[\bar{O}+(1-\theta)\frac{W_{Hjt}}{R_{jt}}\right]+L_{jt}\left[\bar{O}+(1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{15}$$

The equilibrium is defined in appendix section A.4. I also treat existence and uniqueness properties of the model. Finally, I also report a discussion about some assumptions of the model such as the spatial technology diffusion, SBTC and parameter choice.

4 Estimation and Calibration of the Model

The numerical computation of the equilibrium of the model involves recruiting values for all parameters used in the equations above in addition to the values for initial productivity levels, ξ_{kj0}. After obtaining these parameters, I compute the dynamic equilibrium by iterating a system of equations. In order to calibrate the model, I estimate the 10 parameters $\{\theta, \gamma^p, \gamma^L, \gamma^H, \rho, \phi^H, \phi^L, \lambda^H, \lambda^L, \gamma^2\}$ internally within the framework. There are two main reasons why I choose estimation over external calibration for the core parameters. First, using parameters from the literature that studies other periods produces inaccuracies. Second, in order to conduct a quantitative rather than a qualitative analysis, I need to disentangle the quantitative importance of each of the model's parameters. Specifically, I need to distinguish the effect of agglomeration forces from the effects produced by SBTC. Therefore, an identification procedure is necessary to clarify the individual importance of each parameter. I calibrate the residuals parameters $\{m_{2H}, m_{2L}, \mu, \bar{O}, \alpha,\ \mu_d, \forall d\}$ with data from the literature.

4.1 Estimation of the Model

To estimate the set above of 10 parameters and initial ξ_{H0} and ξ_{L0}, I exploit local exposure to SBTC. *It is very important to notice that all the fit of the model with the data and the counterfactuals are conducted with a national change in SBTC, not local.* This aligns with the main point of the paper that is to study how the interaction between a national change in technology and local agglomeration had local implications.

4.1.1 Skill-Biased Technical Change

Autor and Dorn (2013) rank commuting zones by the intensity of the routine occupations.[22] The authors build an index of routinization in which they categorize all occupations by their intensity of routinization. Each occupation υ is defined as routinized if the RTI (or routine task intensity) is higher than the 66th percentile. If an occupation is defined as routinized, the arrival of computers will have a large effect on it because routine occupations and computers are substitutes. For instance, the car industry in Detroit was very affected by skill-biased technology (or routinization, in specific) because the share of laborers working in routine-intensive occupations was very high for both highly and less skilled workers. Using the same approach, I construct the RTI for both highly and less skilled workers in each occupation, as shown in equations 16 and 17.

$$\Delta S_{Ljt} = \sum_{\upsilon=1}^{\Upsilon} \left(\frac{L_{j\upsilon t}}{L_{jt}} - \frac{L_{j\upsilon t-10}}{L_{jt-10}} \right) 1 \left(RTI_{\upsilon} > RTI_{P66} \right) \tag{16}$$

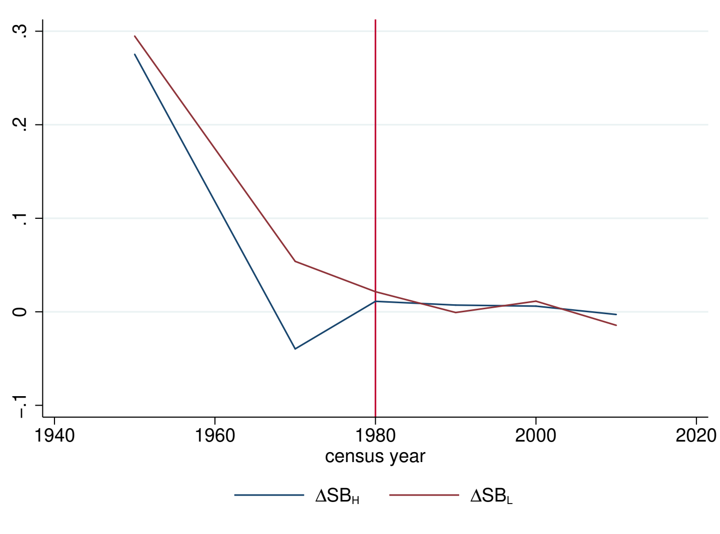
$$\Delta S_{Hjt} = \sum_{\upsilon=1}^{\Upsilon} \left(\frac{H_{j\upsilon t}}{H_{jt}} - \frac{H_{j\upsilon t-10}}{H_{jt-10}} \right) 1 \left(RTI_{\upsilon} > RTI_{P66} \right) \tag{17}$$

Autor and Dorn (2013) find that when the price of computers starts falling, workers in routinized occupations, who are substitutable by computers, see their wages erode. Therefore, MSAs that specialized in routine occupations, both for highly and less skilled workers, experience relative wage declines. ΔS_{kjt} capture this idea well through the measure of routinization. Using this same approach, I build the RTI in each occupation both for the highly skilled and less skilled workers as in equations 16 and 17. ΔS_{Hjt} and ΔS_{Ljt} are two good proxies for how SBTC affects cities in different ways depending on their composition.[23] Figure 6 shows the evolution over time of the aggregate measures. However, this is not a good measure of a productivity shock because it correlates with contemporaneous and local changes that could affect wages. Following the approach of Autor and Dorn (2013), I use national employment changes both for the highly skilled and less skilled workers that are interacted with the share of RTI for the local industry 10 years ago as

[22] For a full definition of commuting zones, refer to the following link from the *United States Department of Agriculture*: http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/

[23] While this approach provides a good proxy for the local impact of SBTC, it might not be the only one. Computer prices might represent the arrival of computers and demonstrate how different cities are affected differently by computer adoption. Beaudry et al. (2010) uses this approach. However, the available data stops in 2000. This shortfall prevents me from recreating the full analysis through 2010 and is insufficient to estimate my model. For this reason, I use Autor and Dorn (2013) approach, which is very flexible with data and allows me to build an index for all years in the analysis.

Figure 6: ΔS_{Hjt} and ΔS_{Ljt} aggregate over time



instruments for ΔS_{Ljt} and ΔS_{Hjt}. These instruments can be described as:

$$\Delta\tilde{S}_{Hjt-10} = \sum_d (H_{d-jt} - H_{d-jt-10})(R_{djt-10}) \quad \text{and} \quad \Delta\tilde{S}_{Ljt-10} = \sum_d (L_{d-jt} - L_{d-jt-10})(R_{djt-10})$$

where $-j$ is all cities in the sample other than MSA j, d is industries in the economy, and t is time. H_{d-jt} and L_{d-jt} are, respectively, the number of highly skilled and less skilled workers in each industry d at the national level at time t that excludes MSA j to avoid mechanical correlations. $H_{d-jt-10}$ and $L_{d-jt-10}$ are the same lagged 10 years. R_{djt-10} is the share of routine occupations among workers in each industry in a specific MSA j. Unlike Autor and Dorn (2013), I create both the index and the instrument for highly skilled H and less skilled L. In this way, I produce extra variation in the data and use the differential impact of technological shocks on the two categories of workers. Differently from Autor and Dorn (2013) that aims at capturing the polarization, here I also want to capture the skill differences. These instrumental variables, $\Delta\hat{S}_{Ljt-10}$ and $\Delta\hat{S}_{Ljt-10}$, are useful in the estimation of the parameters of the model and in the construction of the moment condition.

Table 5 presents the first-stage estimates for these instrumental variables. The predictive relationship between ΔS_H and $\Delta\hat{S}_H$ is sizable and highly significant with F-statistics of 10 or above in each decade as shown in Panel A. The predictive relationship between ΔS and $\Delta\hat{S}_L$ is sizable and highly significant with F-statistics of 10 or above for the decades after 1980. However, the F-statistics for the 1950s, 1970s, and 1980s are less than 10. Specifically, in the 1970s, the F-statistic is less than 7.[24]

4.1.2 Labor Demand

In order to estimate labor demand I use moment conditions that start from the labor demand curves for highly and less skilled workers. The change in productivity levels that interact with changes in demand shocks help to identify the core parameters. Using these conditions, I create moments in order to estimate the set of parameters: $\{\gamma^H, \gamma^L, \phi^H, \phi^L, \rho, \lambda^H, \lambda^L \gamma^2\}$. For this purpose, I start by taking the logs and the first differences of equations 12, 13, which, respectively, give:

$$\Delta w_{Ljt} = (1-\rho)\Delta \text{ln} Y_{djt}(\rho,\, \gamma^{H,}\, \gamma^L,\, \phi^{H,}\, \phi^L) + (\rho-1)\Delta \text{ln} L_{djt} + \gamma^L \Delta \text{ln}\frac{H_{jt}}{L_{jt}} + \\ + \phi^L \Delta \text{ln}\left(H_{jt}+L_{jt}\right) + \lambda^L \Delta S_{Ljt} + \Delta\xi_{Ldjt}$$

$$\Delta w_{Hjt} = (1-\rho)\Delta \text{ln} Y_{djt}(\rho,\, \gamma^H\, \gamma^L,\, \phi_{H,}\, \phi^L) + (\rho-1)\Delta \text{ln} H_{djt} + \gamma^H \Delta \text{ln}\frac{H_{jt}}{L_{jt}} + \\ + \phi^H \Delta \text{ln}\left(H_{jt}+L_{jt}\right) + \lambda^H \Delta S_{Hjt} + \Delta\xi_{Hdjt}$$

Then, by using equation 11 and writing it in first differences, I can isolate $\Delta\epsilon_{kdjt}$, which is the productivity component uncorrelated with technological diffusion:

$$\Delta\xi_{Hdjt} = \xi_{Hdjt-1}^{\gamma_2}\left(\int_s \xi_{Hdst-1} ds\right)^{1-\gamma_2} - \xi_{Hdjt-2}^{\gamma_2}\left(\int_s \xi_{Hdst-2} ds\right)^{1-\gamma_2} + \Delta\epsilon_{Hdjt} \tag{18}$$

$$\Delta\xi_{Ldjt} = \xi_{Ldjt-1}^{\gamma_2}\left(\int_s \xi_{Ldst-1} ds\right)^{1-\gamma_2} - \xi_{Ldjt-2}^{\gamma_2}\left(\int_s \xi_{Ldst-2} ds\right)^{1-\gamma_2} + \Delta\epsilon_{Ldjt} \tag{19}$$

Therefore, we can replace 18 and 19 respectively into 12 and 13 isolating $\Delta\epsilon_{Hdjt}$ and $\Delta\epsilon_{Ldjt}$. In the spirit of Diamond (2016) and Suárez Serrato and Zidar (2016), the identification strategy follows from the changes in the labor supply that are uncorrelated with local productivity. Differently from previous work, this estimation strategy takes into account the dynamic component of the model and the fact that current local productivity is correlated with past productivity changes. Also, the interaction of SBTC shocks with the cities' housing supply elasticities leads to variation in the labor supply that is uncorrelated with the unobserved changes in local productivity. The housing supply affects the migration decisions in response to a labor demand shock. Differential housing

[24] As a robustness test, I estimate the model without the 1950s and the parameter estimates are unchanged.

supply elasticities generate exogenous variation in the labor supply. I compare two cities, one has a very elastic housing supply and the other has a very inelastic one, both experience an increase in labor demand; and workers move to take advantage of these increases. But, once they move, the MSA with more inelastic housing will have a higher increase in housing prices. Therefore, the rent increase will prevent more in-migration in the MSA with higher housing prices for the same level of labor demand shock that offsets the increase in wage through the labor-demand channel. Specifically, the exclusion restrictions are:[25]

$$\mathrm{E}(\Delta\epsilon_{Hdjt}\Delta Z_{jt}) = 0 \quad \text{and} \quad \mathrm{E}(\Delta\epsilon_{Ldjt}\Delta Z_{jt}) = 0$$

$$\text{Instruments} : \Delta Z_{jt} = \begin{pmatrix} \Delta\hat{S}_{Ljt}x_j^{reg} & \Delta\hat{S}_{Hjt}x_j^{reg} \\ \Delta\hat{S}_{Ljt}x_j^{unav} & \Delta\hat{S}_{Hjt}x_j^{unav} \end{pmatrix}$$

The moment conditions are jointly combined with identifying cities' supply curves and workers' labor supply to cities. Finally, they will be jointly estimated with a two-step GMM procedure.

4.1.3 Labor Supply

As specified above, the indirect utility for agent i of type k living in MSA j at time t can be written as

$$V_{ikjt} = \delta_{kjt} + \zeta_{ijt}$$

where

$$\begin{aligned} \delta_{kjt} = \Bigg[& \theta \ln(W_{kjt} - R_{jt}\bar{O}) + \\ & +(1-\theta)\left[\ln(1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O})\right] + (1-\theta)\left[\ln((1-\theta)W_{kjt} - R_{jt}\bar{O})\right] + \\ & +\gamma^p \ln\left(H_{jt}/L_{jt}\right) + A_{kjt}\Bigg] \end{aligned}$$

The fact that the model does not rely on the agents' history simplifies the estimation procedure by causing it to resemble a static framework. The estimation of the labor supply follows from the decision of the agents on where to live in each period. Because the utility component δ_{kjt} does not depend on individual worker characteristics, the estimates for each type k is exactly equal to

[25] To improve the estimation, I supplement the routinization shock with a "classic" Bartik instrument. This instrument increases the precision of the estimators.

the ln population of each demographic group observed living in the MSA. Therefore, this is a simplification with respect to Berry et al. (2004). I take the difference in mean utility δ_{kjt} over time to get:

$$\Delta\delta_{kjt} = \theta \ln \frac{W_{kjt} - R_{jt}\bar{O}}{(W_{kjt-10} - R_{jt-10}\bar{O})} +$$

$$+(1-\theta)\frac{\ln(1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}}{\ln(1-\theta)\frac{W_{kjt-10}}{R_{jt-10}} + \bar{O}} + (1-\theta)\frac{\ln((1-\theta)W_{kjt} - R_{jt}\bar{O})}{\ln((1-\theta)W_{kjt-10} - R_{jt-10}\bar{O})} +$$

$$+\gamma^p \ln \Delta(H_{jt}/L_{jt}) + \Delta A_{kjt}$$

Identifying workers' preferences for wages, rent, non-traded local goods, housing, and amenities requires variation in these MSA characteristics that is uncorrelated with local unobservable amenities ΔA_{kjt}. This reasoning follows Diamond (2016). Specifically, I use SBTC shocks and their interaction with the characteristics of the supply elasticity. For the exclusion restriction to be satisfied, the set of instruments needs to be uncorrelated with unobserved exogenous changes in the MSA's local amenities. The key idea is that since SBTC shocks are driven by national changes in industrial productivity, these shocks are unrelated to changes in local exogenous amenities. These instruments can be supplemented with data to provide extra power in the identification process. Specifically, I obtain the share of household expenditure on non-tradable goods, θ, from the literature. I also estimate the model without using the externally calibrated data by relying only on the instruments for identification. Specifically, the moment restrictions are:

$$\mathrm{E}(\Delta A_{Hjt}\Delta Z_{jt}) = 0 \qquad \text{and} \qquad \mathrm{E}(\Delta A_{Ljt}\Delta Z_{jt}) = 0$$

$$\text{Instruments: } \Delta Z_{jt} = \begin{pmatrix} \Delta\hat{S}_{Ljt} & \Delta\hat{S}_{Hjt} \\ \Delta\hat{S}_{Ljt}x_j^{reg} & \Delta\hat{S}_{Hjt}x_j^{reg} \\ \Delta\hat{S}_{Ljt}x_j^{unav} & \Delta\hat{S}_{Hjt}x_j^{unav} \end{pmatrix}$$

All parameters are jointly estimated in a 2-stage GMM where standard errors are clustered at the MSA level and there are decade fixed effects to account for national changes. Further, I test whether the over-identification restrictions can be jointly satisfied.[26]

[26] To improve the precision of the estimates, I also add standard Bartik shocks in the instrumental set as in Diamond

4.2 Migration Costs

By taking the differences in δ_{kjt}, migration costs $m_{k2}(j)$ are eliminated since they do not vary over time. Therefore, another strategy is needed to calibrate the migration costs. To do so, I rely on existing literature that has estimated migration costs using structural models. A famous work from Kennan and Walker (2011) provides estimates of large moving costs from one state to the other in the US but does not distinguish between high and less skilled workers. Notowidigdo (2011), instead, provides separate migration costs for highly and less skilled workers, closely related to what the migration costs in my settings look like. This makes his estimates applicable to my setting. Specifically, Notowidigdo (2011) uses an exponential function to estimate migration costs. The functional form he estimates is as follows:

$$m_{k2}(j) = \frac{\sigma^k exp(\beta^k x_j) - 1}{\beta^k}$$

where x_j relates to MSA characteristics such as population. This functional form is very flexible since, despite having only two parameters, it has advantageous curvature features as Notowidigdo (2011) discusses. Specifically, migration costs increase a lot when characteristics such as population are low and increase rather than at higher levels of population. In order to check how sensitive my model is to the estimates of the migration costs, I run some sensitivity analysis. The results do change only mildly.

4.2.1 Estimation Results and Robustness

I use a GMM estimation procedure with data at the MSA level for 1940-2010 with data every 10 years and 14 industries. The results are reported in Table 6 where I run five model specifications to test the sensitivity of the model. The main differences across the specifications are the endogenous productivity spillover of size and skill ratio and the endogenous amenity supply. The first specification only includes the endogenous spillover effect of skill. The second only include also the spillover effects on population size. The third specification adds the endogenous amenities but removes the spillover effects on population size. The fourth specification reports a "classic" model with no production spillover, but only with endogenous amenities. The fifth and last specification reports the full model that will be used for the rest of the analysis below. The results of the estimates, overall, are thus in accordance with the literature.

(2016).

Panel A reports the estimates for the labor supply parameters. Specifically, the share of expenditure in housing, θ and the preferences for endogenous amenities γ^p. In column (1) and in column (2), the estimates of θ is between 0.618 and 0.608. This indicates that about 60% of the expenditures are on housing relative to the tradable good *T*. In column (3)-(5), when we include the endogenous amenities, the estimates drop to a range between 0.421 and 0.446. This suggests that the preference on housing shrinks when individuals assign a value to the local amenities. The second parameter of interest of the labor supply side is γ^p estimated in specifications (3)-(5). In the three specifications, the estimates are very similar. A 1% increase in the skill ratio increases the local highly and less skilled working population by 0.927% in column (5). This value is lower than that in Diamond (2016) and Albouy (2012). One explanation could be that my estimates relate to a longer time period. Overall, these estimates show that workers, in general, prefer cities with higher wages, lower rents, higher college share.

Panel B, instead, reports the parameters for the labor demand such as the inverse of the elasticity of substitution ρ; the spillover effects of skill, γ^H and γ^L; the spillover effects of population size, ϕ^H and ϕ^L; the scale effects on SB^H and SB^L. Additionally, it also reports the estimate on γ^2, the spatial technology diffusion parameter. In column (1), in the model with spillover effect of skills, I estimate an elasticity of substitution of 2.5, which is the same as in Card (2009). This elasticity jumps to 3.9 when I include also the spillover effects on size. In column (3), where I remove the spillover effects on size but we do add the endogenous amenities, then, the estimate is 2.1, closer to column (1). In the model with no spillover effects at all, the estimates drops to 1.47. Finally, column (5), where I estimate the full model, the elasticity of substitution is 3.03. Overall, the estimates seem to suggest that in a model with the spillover effects on production rather than the "classic" model, the elasticity of substitution between higher and less skilled workers increases. The estimates of γ^H in all the specifications is positive and statistically significant. In the specification of the model with no spillover effects of size (column (1) and column (3)), a 1% increase in the share of highly skilled workers raises their wages by 0.72%. Instead, in column (2) and (5), that include spillover effects of size, a 1% increase in the share of highly skilled workers raises their wages by 0.545% and .528%, respectively. My estimates of γ^H are higher than the one in Diamond (2016) and Moretti (2004) that find an estimate of 0.322 and 0.16, respectively. This suggests that part of the positive effect of size is captured by skill composition. Overall, having a higher share of highly skilled workers increases highly skilled workers' wages by a non-negligible amount. The same cannot be told about the wages of less skilled workers. In fact, the estimates of

γ^L are between 0.056 and 0.118 but they are not statistically significant in any of the specifications. The estimates of the spillover effects of size, ϕ^H for highly skilled workers are 0.271 and 0.253 in column (2) and (5), respectively. This is equivalent to say that in the full model, a 1% increase in the size of the city's workforce raises the highly skilled wages by 0.253%. This effect is smaller than the skill ratio effect but it is still non-negligible. For less skilled workers, the same doesn't hold. Baum-Snow and Pavan (2013) estimate that at least 23% of the overall increase in the variance of log hourly wages in the US from 1979 to 2007 is explained by the more rapid growth in the variance of log wages in larger locations relative to smaller locations. The estimates in column (2) and (5) are around 0.210 and 0.211 but they are not statistically significant. Overall, the estimates of the production spillover seem to suggest that being in a larger city and in a more skilled city increases the wages of highly skilled workers but does not affect the wages of less skilled workers. The estimates also suggest that the spillover on the skill ratio is quantitatively more important than the size spillover. I also report the estimate for the coefficient for the SBTC measure. This estimate serves as a scale of the effect for the rest of the analysis. λ^H, the estimate on the highly skilled are slightly negative and close to 0. Instead, the estimates on λ^L are strictly negative and statistically significant ranging between -.160 and -.998. This suggests that routinization had a negative impact on the wages of the lower skilled workers but not on highly skilled workers.

Panel C reports the estimate of γ^2 which is the most important parameter to regulate the degree of spatial technology diffusion. With the except of specification (4) that does not have spillover effects, the estimates are between 0.934 and 0.998. In the full model, we can interpret the result as saying that technology diffuses at a rate of 4.9% every 10 years. If we compare this estimate to Desmet et al. (2018), that use a value of 0.99 per year, then my estimates imply slightly lover technology diffusion.

4.2.2 Other Calibrated Parameters

To complete the calibration of the model and compute its equilibrium, I borrow the other parameters from the literature. These values are reported in Table 7. To include housing in the model with non-homothetic preferences, I also include a subsistence level of housing, $\bar{O}$, from Ganong and Shoag (2017), which is set to match the Engel curve for housing. To complete the housing sector, I estimate a value for the elasticity of housing, μ. This elasticity is also borrowed from Ganong and Shoag (2017). I chose this elasticity to generate a one-to-one relationship between log prices and log per capita incomes in order to match the relationship from the data. The elasticity is equal to 0.4. This parameter decreases to 0.135 for the cities with higher regulations after 1980. The parameters of the migration cost function, which is exponential, are different for highly skilled and less skilled workers. I borrow these estimates from Notowidigdo (2011), which uses an identification strategy based on Bartik instruments. Given this functional form, it turns out that the migration costs are about 1.16 that is higher for less skilled workers than for highly skilled workers whose costs are equal to one. Another set of migration costs could have been estimated by using the same approach as in Desmet et al. (2018). While Desmet et al. (2018) uses this procedure for one type of workers, the analysis could be extended to two types of workers.[27]

5 Model Simulation and Counterfactuals

In this section, I first provide more details as to how I achieve the numerical computation of the equilibrium. Second, I show how the model matches the non-targeted moments in the decline in β-convergence. Specifically, the model fits well the decline in spatial convergence for highly skilled workers. Third, I conduct a quantitative decomposition of each mechanism's effect on the decline in convergence. Fourth, I investigate whether the model matches other non-targeted moments such as the "The Great Divergence" of skills, the secular decline in migration, and the increase in wage dispersion among others. Fifth, I show how the increase in regional inequality fostered the trade-off between aggregate inequality and growth.

The estimation procedure obtains the values for all 10 model parameters, the initial productivity terms, and national measures of SBTC. Next, I compute the equilibrium of the model by solving a system of equations for every period t that incorporates the productivity values from the previous period.

The model can be reduced to 46 equations, as shown in the appendix. Given that the analysis includes 240 cities, the iteration procedure contains 11,040 equations for each period t. The equilibrium conditions correspond to equations 7, 8, 12 and 13. Because of the large number of cities, the problem is highly dimensional. An extra complication to the model is the endogenous agglomeration effects that could induce the system of equations to explode. However, the estimates respect the restrictions imposed by the system and are stable. As a robustness test, I conduct a sensitivity analysis and check whether the variation in the parameters changes the results substantially and whether the system maintains wage convergence. More details about these conditions can be

[27] Extending the migration cost algorithm is not the primary focus of this paper and, therefore, it is left for future work. The sensitivity analysis indicates that the migration costs do not change the non-targeted moments much.

found in appendix's section D.1.

5.1 Model vs. Data

In this section, I test how the model fits the data on regional convergence as shown in figure 1 and 2. Practically, I proceed in the following way:

1. I obtain the parameters estimates and ξ_{H0} and ξ_{L0} from the GMM estimation;
2. I initialize the model for 1940 using those values and a *national* average of S_{Ht} and S_{Lt};
3. I solve the model for every period until 2010;
4. With the solved model, I run the same regression within my model as I did with the data in section 2. Specifically, with model-generated measures of average wages, wages for highly and less skilled workers since 1960, I estimate in the model the β, β^k-convergence using the specification 1 proposed in section 2, which follows from Baumol (1986).

In Figure 7, I plot the estimated β-convergence from the model and from the data to compare them. The x-axis of figure 7 report the initial year of the convergence equation 1, therefore, the analysis covers both the period since 1960 Overall, the match is satisfactory. The estimates from the data and the model differ only by 0.005% points.

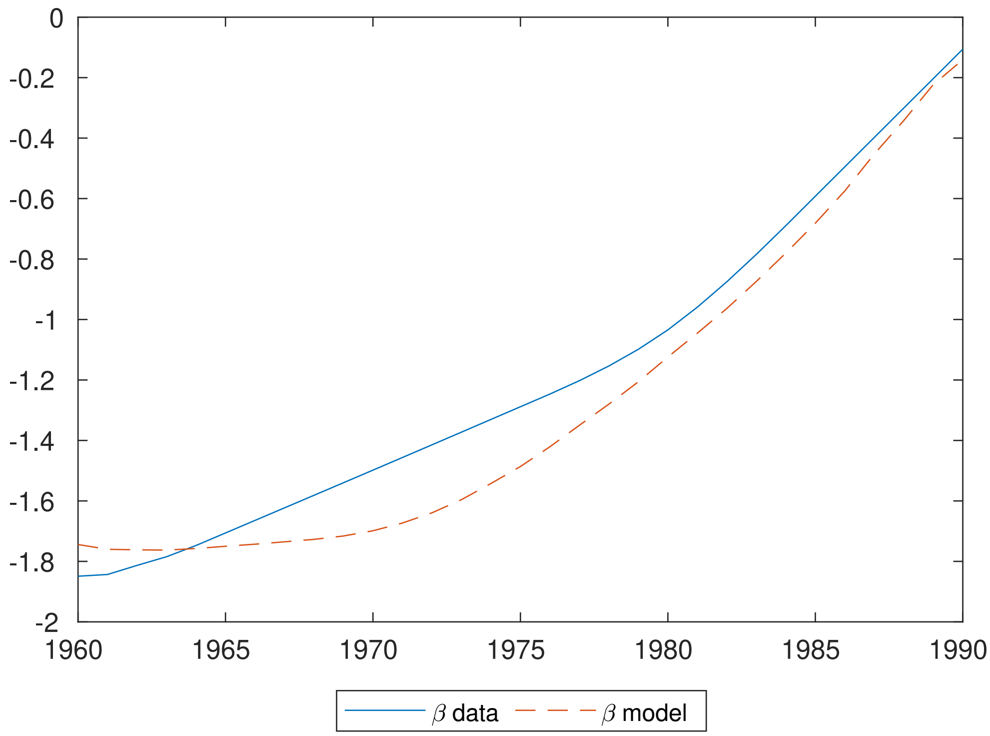
The left (right) plot in figure 8 compares the $\beta^H(\beta^L)$-convergence rates over time both in the data and in the model. The estimates are very close over time. The model performs very satisfactorily in fitting the wage convergence patterns in the non-targeted moment and the decline of convergence for the higher (less)-skill group.

5.2 Quantitative Decomposition

After ensuring that the model fits the data, I calculate several counterfactual scenarios for the β, β^H, and β^L convergence rates to assess the quantitative contributions of each of the model's mechanisms. Specifically, I proceed step-wise and sequentially "turn off" each component of the model that contributes to the decline in wage convergence over time.

My counterfactual of interest is comparing estimates of β, β^H and β^L over time in the baseline model with the estimates that I obtain once I "turn off" the mechanisms step-wise after 1980. Figure 9 shows the 20-year window β convergence starting with 1979 as first year, When I set the agglomeration forces, γ^H, γ^L, ϕ^H, ϕ^L, to 0 starting in year 1980, the change in β is quite dramatic

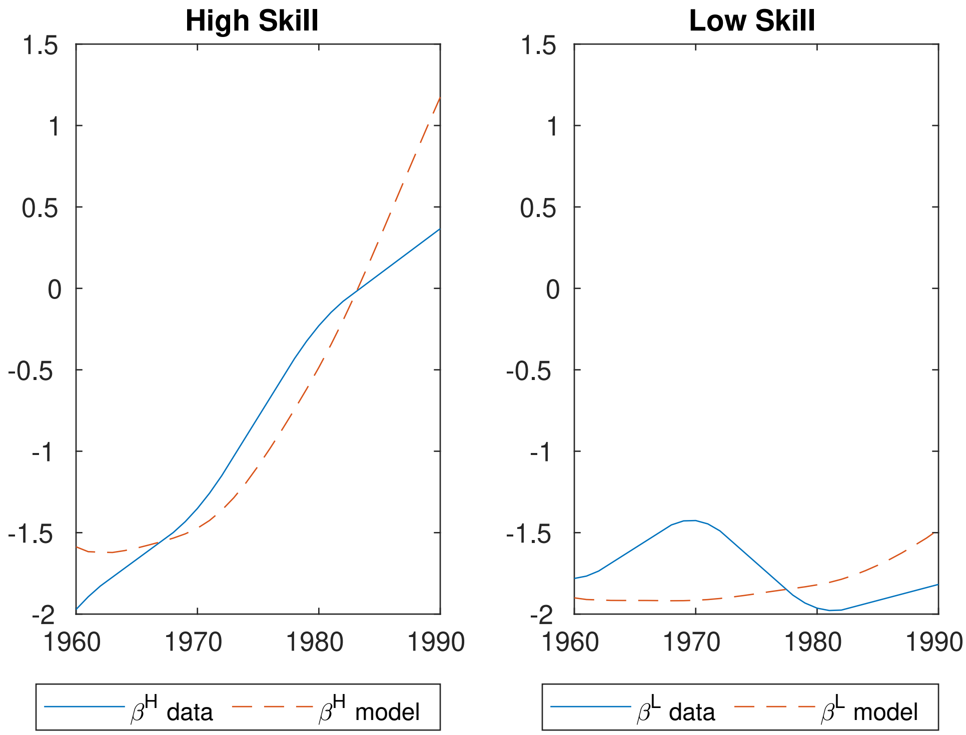
Figure 7: Model Matching the Data on Wage Convergence



Note: This figure shows a rolling estimate of the β-convergence with a rolling window of 20 years. The solid line is the data for which we have observations every 10 years (that I smooth over time), while the dashed line is the estimate of the β-convergence from the model for which we can compute a yearly estimate.

suggesting. Without agglomeration, convergence would have been much higher than it was in the data until 1980-2000 time period. Then, it would have increased substantially initially but it would have set around -0.5% in 1990-2010. When I additionally turn off also SBTC, setting λ^H and λ^L to 0 the model predicts that overall β would have decreased less on impact than accounting for agglomeration forces alone, but in 1990-2010, convergence would have been -1%, so much higher than what is in the data. When I, respectively, set migration cost, m_{H2} and m_{L2}, to 0 after 1980 and housing elasticity η to be the same as in the previous period, and $\bar{O}$ to 0. I find that they move the convergence rate by only a few decimals in percentage points. Overall, the largest decline is explained by the interaction between agglomeration and SBTC. In figure 10, I run the same exercise as above, just isolating how β^H changes over time in the different scenarios on the left plot. The model with no agglomeration shows a similar pattern to the one of figure 9 with an overall decrease in convergence ending at -0.75% in 2010. When turning off also SBTC forces, the overall convergence rate would have declined to -0.78%. When switching off migration costs

Figure 8: Model Matching the Data on High and Less Skilled Wage Convergence

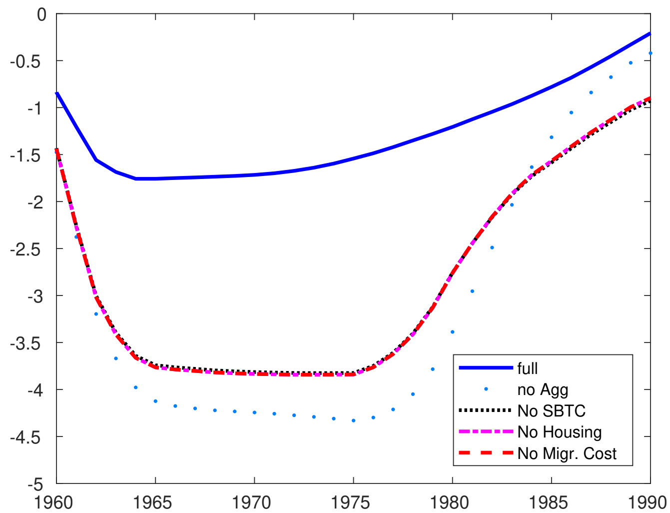


Note: This figure on the left (right) shows a rolling estimate of the $\beta^H (\beta^L)$-convergence with a rolling window of 20 years. The solid line is the data for which I smooth the 20-year rolling estimate, while the dashed line is the estimate of the β-convergence from the model for which I compute a yearly estimate.

and housing regulation changes, respectively, I find that they move the convergence rate by only a few percentage points as for β. On the right plot of figure 10, I run the same exercise for β^L. I find that if agglomeration forces had been set to 0, then, overall, β^L would have decreased to almost -3.5% and finalized to -2.5% in 2010. However, when I also shut down also SBTC forces, the exercise shows that convergence in the last 10-year window would have decreased more than in the no agglomeration case than in the baseline, approximately to -1.7%, due to general equilibrium forces. When turning off housing regulation changes and migration costs, respectively, I find that they move the convergence rate by only few percentage points as for β and β^H. The β-convergence between 1980 and 2010 is estimated to be -1.5% a year. Overall, the main finding is that the bulk of the decline in convergence after 1980 can be attributed to the interaction between SBTC and the agglomeration forces.

The main takeaway of this counterfactual analysis is that the convergence rate with endogenous productivity channels and SBTC are about 1.2% a year. But, if I shut down the productivity

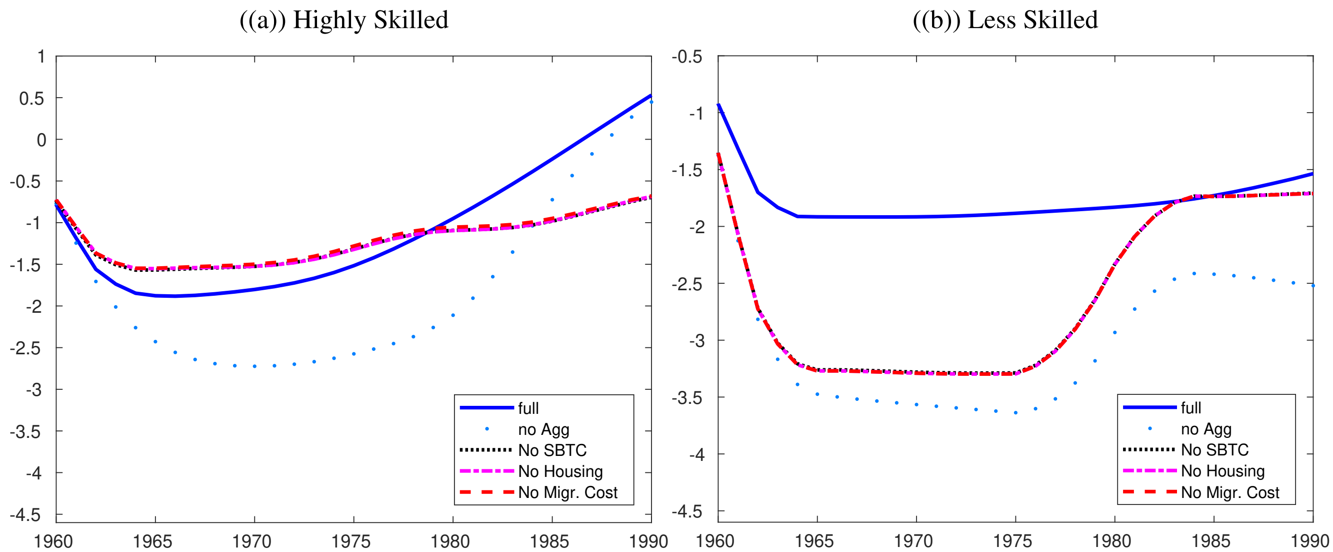
Figure 9: Quantitative Decomposition of Wage Convergence β



Note: This figure shows the counterfactual exercises in which I turn off the agglomeration forces (dotted blue), SBTC (small dotted black), migration cost (alternate dotted pink), and housing (large dotted red) one after another.

channel, nominal wage convergence is about 1.1% a year. Table 9 reports the changes of the β, β^H and β^L estimates with 30-years rolling windows between the 1960-1990 and 1980-2010 in the full model and in the counterfactuals. Each column reports the estimates for a version of the model. Panel A shuts down the element from the full model at the time. Instead, panel B removes elements sequentially. A comparison between panel A and B tell us whether ordering matters. Column (1) reports the estimated for the full model. It suggests that the β estimate change by 100% but most of the change is due to the change in β^H and only mildly due to changes in β^L. Column (2) reports the estimates for the model without agglomeration forces. In this case the β estimate would have dropped only by 17% mostly due to a drop in convergence in β^H while β^L would have even increased. In column (3) without SBTC, the overall convergence would have decreased by 34%, mostly due to a decrease in β^H convergence. Finally, column (4) and (5), respectively, show the results for the model without housing changes and for the model without migration costs. In these two last cases the β estimates change only mildly, but quantitatively are not very important. Column (1) and (2) in panel B are the same as in panel A. In column (3), the change in β estimate would have been only 47% but it is mostly because there would have been a decrease of 74% in β^H. As in panel A, column (4) and (5) only count marginally.

Figure 10: Quantitative Decomposition of Highly and Less Skilled Wage Convergence



Note: This figure shows the counterfactual exercises in which I turn off the agglomeration forces (dotted blue), SBTC (small dotted black), migration cost (alternate dotted pink), and housing (large dotted red) one after another.

5.3 Decline in Gross Migration Flows over time

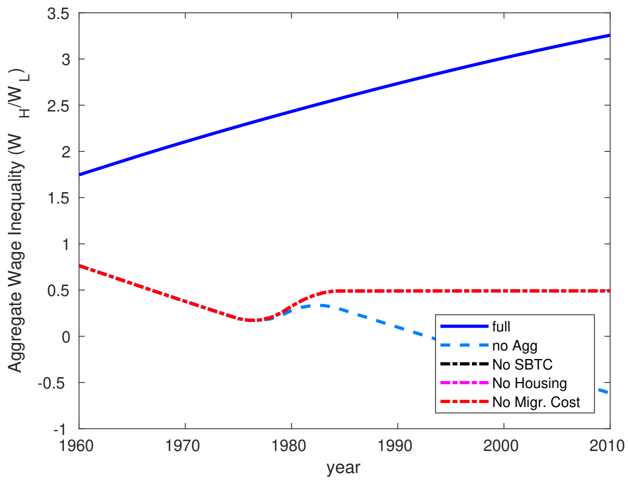
The decline in geographic migration, documented by Molloy et al. (2014), is another important structural change that happened in the US in the last several years. In the early 1990s, about 3% of Americans moved between states each year. But, today that rate has fallen by half. Gross flows of people have declined by around 50% over the last 20 years. Schulhofer-Wohl and Kaplan (2017) provide and test a theory of reduction in the geographic specificity of occupations coupled with information technology and inexpensive travel. They find that these two mechanisms together can explain at least half of the decline in gross migration since 1991. Can my framework help to explain the decline in gross migration flow? Technological innovation increases the sorting of skilled workers into skilled cities, and once workers are sorted, their incentive to move will decrease over time since it will be harder to find a city with similar wages. If, moreover, the technological shock persists over time, then this effect will become even stronger by decreasing migration even further. For instance, suppose that a highly skilled worker lived in San Francisco in the 1980s. When the technology shock arrives, the highly skilled worker would have less incentive to move because San Francisco would have the highest wages for him or her. Another highly skilled worker, who currently lives in Detroit, decides to move to San Francisco. Over time, the incentive to migrate decreases because the workers will have a better match in their current MSA.

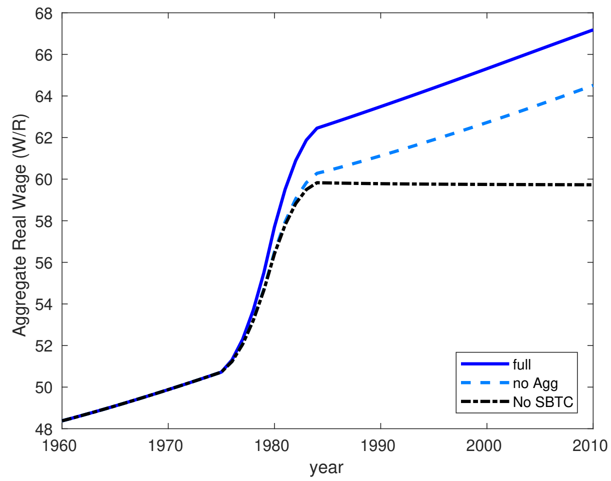
This is supported by the evidence that the migration rate for skilled workers decreased more than the migration rate for less skilled workers. Figure 13 shows that the model matches the data for the trends in the migration rate over time. It shows that the rate of the decline is similar both for the less skilled both in model and data.

5.4 Real Wage Convergence, Aggregate Prices, Inequality and Growth

With data on wages and local prices generated by the model, I calculate the real wages in each city by dividing the nominal wage by the price index. In figure 15, I plot the evolution of the β estimate for real wages. Unfortunately, while the availability of historical data on local prices is very limited, this model allows to produce a time series of prices back to the 1940 by city. The slope of the decline reproduced by the model is similar to the one in the data as the left plot shows. The majority of the decline follows from the decline in the convergence of the highly skilled real wages. The real wage convergence for less skilled workers is still around 2% a year. These findings suggest that the decline in wage convergence did not solely happen for nominal wages but also for real wages, which gives a better sense of the local purchasing power. It could have been the case that prices had adjusted across space such that the differences had shrank, however, both data and model suggest that it is not the case. Figure 14 shows the evolution of the variance of rents and prices across cities. For both sets of prices, the variance was decreasing before 1980, but around that time the trend reverted. So, despite the increase in price heterogeneity across cities, overall, it is still the case that prices are following the trends in wages. Finally, using the estimates of wages and prices at the local level, we can use the model to quantify the nationally aggregate real wage growth predicted by the model. Then, we ask the question on how much the national real wage growth would have been if the national SBTC shock had not happened and had not had an effect through the mechanism presented in this paper. The left panel in figure 11 shows the time evolution of skill premium. The model reproduces an increase in the national skill premium as we see in the data with a jump around 1980. However, if agglomeration forces had not been there and SBTC neither, then, we would have not observed an increase in the skill premium. The right panel of figure 11, instead, shows the time evolution of the national real wages. The model suggests that real wages growth would have been about .83% a year between 1960 and 2010, however, if SBTC had not happened, aggregate growth would have been only .51% a year in the same time frame. Therefore, if on the one hand, SBTC induced more inequality both at the national and regional level, at the same time, it increased overall aggregate growth by the sorting mechanism of highly

Figure 11: Time Evolution of Aggregate Skill Premium and Real Wages





skilled workers to highly skilled cities.

6 Conclusions and Future Directions

In this paper, I show that the decline in wage convergence among MSAs that is observed after 1980 is largely due to the decline in wage convergence among highly skilled workers, whereas the wage convergence among less skilled workers does not decline at all. Thus, any account of convergence and its decline must distinguish between skill groups. Motivated by this observation, I develop a quantitative spatial equilibrium model with technology diffusion to reconcile the patterns of regional convergence and other salient moments' changes in the regional data in the last 70 years; and determine the aggregate growth and inequality implications of those. To the best of my knowledge, this paper is the first to point to the interaction of technology and space in the trade-off between inequality and growth.

Understanding what stopped income convergence across the US regions and increased income inequality for different levels of skills might have important policy implications especially for the regions that are not able to grow like richer regions. Dealing with sustaining the growth in the richest MSAs and arresting the decline in poorer MSAs is an important challenge for policymakers.

This paper plants the seeds for a broader research agenda. Specifically, how can we rationalize the regional inequality patterns across countries? The framework of this paper is flexible enough that it can be extended to perform several types of analysis, including a cross-within-country analysis. In recent work, we find that the speed of regional convergence has increased across countries but it has decreased within country for a large sample of countries in the world. In a nutshell, although poorer countries keep growing at a faster rate than rich countries, their growth rates are mainly driven by few regions, leaving others behind (Chatterjee and Giannone, 2021). This highlights the pressure to deepen our understanding of the changing patterns of regional convergence.

References

Acemoglu, Daron and David Autor, *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Vol. 4 of *Handbook of Labor Economics*, Elsevier,

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," Working Paper 14472, National Bureau of Economic Research November 2012.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Autor, David and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market," *American Economic Review*, 2013, (103(5), 1553–1597).

Autor, David H., David Dorn, and Gordon H. Hanson, "Untangling Trade and Technology: Evidence from Local Labour Markets," *The Economic Journal*, 2015, *125* (584), 621–646.

Bajona, Claustre and Timothy Kehoe, "Trade, Growth, and Convergence in a Dynamic Heckscher-Ohlin Model," *Review of Economic Dynamics*, July 2010, *13* (3), 487–513.

Barro, Robert J. and Xavier Sala i Martin, "Convergence," *Journal of Political Economy*, 1992, *100* (2), 223–251.

_ and _ , "Technological Diffusion, Convergence, and Growth," *Journal of Economic Growth*, Mar 1997, *2* (1), 1–26.

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

_ , Matthew Freedman, and Ronni Pavan, "Why Has Urban Inequality Increased?," *American Economic Journal: Applied Economics*, Forthcoming.

Baumol, William, "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show," *American Economic Review*, 1986, *76* (5), 1072–85.

Beaudry, Paul, Mark Doms, and Ethan Lewis, "Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 2010, *118* (5), 988 – 1036.

Bernard, Andrew B and Charles I Jones, "Productivity and Convergence across U.S. States and Industries," *Empirical Economics*, 1996, *21* (1), 113–35.

Berry, Christopher R. and Edward L. Glaeser, "The Divergence of Human Capital Levels Across Cities," September 2005, (11617).

Berry, Steven, James Levinsohn, and Ariel Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, February 2004, *112* (1), 68–105.

Bound, John and George Johnson, "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *American Economic Review*, June 1992, *82* (3), 371–92.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the U.S.," Working Paper 23330, National Bureau of Economic Research April 2017.

Caliendo, Lorenzo, "On the Dynamics of the Heckscher-Ohlin Theory," 2011.

_ **, Maximiliano Dvorkin, and Fernando Parro**, "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Card, David, "Immigration and inequality," *American Economic Review*, 2009, *99* (2), 1–21.

_ **and John E. DiNardo**, "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, October 2002, *20* (4), 733–783.

Caselli, Francesco and Wilbur John Coleman, "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, June 2001, *109* (3), 584–616.

Chatterjee, Shoumitro and Elisa Giannone, "Unequal Global Convergence," 2021.

Comin, Diego A. and Martí Mestieri Ferrer, "If Technology Has Arrived Everywhere, Why has Income Diverged?," *American Economic Journal: Macroeconomics*, Forthcoming, (19010).

Davis, Donald R. and Jonathan I. Dingel, "The Comparative Advantage of Cities," *NBER Working Papers*, October 2014.

Desmet, Klaus and Esteban Rossi-Hansberg, "Spatial Development," *American Economic Review*, April 2014, *104* (4), 1211–43.

_ **, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg**, "The Geography of Development," *Journal of Political Economy*, 2018, *126* (3), 903–983.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, March 2016, *106* (3), 479–524.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh, "Skilled Scalable Services: The New Urban Bias in Economic Growth," 2020.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny, "Spatial Sorting," *Journal of Political Economy*, 2014, *122* (3), 554–620.

Faberman, R. Jason and Matthew Freedman, "The urban density premium across establishments," *Journal of Urban Economics*, 2016, *93* (C), 71–84.

Fajgelbaum, Pablo and Cecile Gaubert, "Optimal Spatial Policies, Geography and Sorting," Working Paper 24632, National Bureau of Economic Research May 2018.

Feenstra, Robert and Gordon Hanson, "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates For the United States, 1979–1990," *The Quarterly Journal of Economics*, 1999, *114* (3), 907–940.

Ganong, Peter and Daniel Shoag, "Why Has Regional Convergence in the U.S. Stopped?," *Journal of Urban Economics*, June 2017, *102*, 76–90.

Gennaioli, Nicola, Rafael LaPorta, Florencio Lopez de Silanes, and Andrei Shleifer, "Growth in Regions," *Journal of Economic Growth*, 2014, *19* (3), 259–309.

Glaeser, Edward L. and Joseph Gyourko, "Housing Dynamics," December 2006, (12787).

Higgins, Matthew, Daniel Levy, and Andrew Young, "Growth and Convergence across the United States: Evidence from County-Level Data," *The Review of Economics and Statistics*, 2006, *88* (4), 671–681.

Holmes, Thomas J., "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, August 1998, *106* (4), 667–705.

Hsieh, Chang-Tai and Enrico Moretti, "Why Do Cities Matter? Local Growth and Aggregate Growth," May 2015, (21154).

_ , Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," January 2013, (18693).

Katz, Lawrence F. and Kevin M. Murphy, "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 1992, *107* (1), 35–78.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Krusell, Per, Lee E. Ohanian, JosÈ-Victor RÌos-Rull, and Giovanni L. Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, September 2000, *68* (5), 1029–1054.

Levy, Frank and Richard J Murnane, "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, September 1992, *30* (3), 1333–81.

Lyon, Spencer and Michael E Waugh, "Quantifying the losses from international trade," *Unpublished manuscript*, 2019.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behaviour," in P. Zarembka, ed., *Frontiers in Econometrics*, New York, NY, USA: Academic Press New York, 1973, pp. 105–142.

Molloy, Raven, Christopher L. Smith, and Abigail K. Wozniak, "Declining Migration within the U.S.: The Role of the Labor Market," April 2014, (20065).

Moretti, Enrico, "Human Capital Externalities in Cities," *Handbook of Regional and Urban Economics*, 2004.

_ , *The New Geography of Jobs*, Houghton Mifflin Harcourt, 2012.

Nagy, David, "City location and economic development," 2016 Meeting Papers 307, Society for Economic Dynamics 2016.

Notowidigdo, Matthew J., "The Incidence of Local Labor Demand Shocks," June 2011, (17167).

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 1982, *90* (6), 1257–1278.

Rosen, Susan, "Wages-based Indexes of Urban Quality of Life," *Current Issues in Urban Economics*, 1979.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, , and Matthew Sobek, "Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]," 2015.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Schulhofer-Wohl, Sam and Greg Kaplan, "Understanding the Long-Run Decline in Interstate Migration: Online Appendix," *International Economic Review*, 2017, *58* (1), 57–94.

Serrato, Juan Carlos Suárez and Owen Zidar, "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, September 2016, *106* (9), 2582–2624.

A Appendix

A.1 Definitions

MSA The unit of geography is the metropolitan statistical area (MSA) that is "a region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core." I rank the MSAs by share of highly skilled workers over less skilled workers. I define "highly skilled" MSAs as those that have a concentration of highly skilled workers larger than the national average. The remainder are defined as "less skilled" MSAs. I refer to MSAs as cities in the first part of the paper for a less technical discussion. There are two main reasons why I pick MSAs over states or over counties. First, MSAs are the smallest unit of analysis for which I can measure wages by skill group, number of highly and less skilled, rent by skill group back to 1940. Second, MSAs are consistent with the mechanism I want to explain in this paper. For instance, agglomeration happens in San Francisco, not in California. The Census consistently includes 240 MSAs across all four decades from 1980 to 2010 but from 1940 to 1970. Following the definitions of metropolitan and micropolitan statistical areas, I try to homogenize the definitions of MSAs over time. However, this is not possible for all cites.[28]

Highly and Less skilled Workers I follow the previous work such as Acemoglu and Autor (2011) that use education as a proxy for skills. Then, I create two groups: "highly skilled" workers are the ones who have at least a 4-year bachelor's degree while "less skilled" workers are those whose education is less than that.

Composition Adjusted Wages I compute hourly wages at the individual level as annual wages divided by the number of hours worked in the last year. My estimation sample consists of individuals between 21 and 55 years of age who were employed at least 40 weeks per year and were not self-employed. However, for a robustness check, I relax the sample restrictions and, qualitatively, the results are unchanged. To conduct my analysis, I do a compositional adjustment to the wage measure reported in the Census data. This is possible thanks to the high dimensionality of the available data. I adjust the wages for age, sex, nativity, and race. The changing composition of workers could explain some of the variation in nominal wages across MSAs over time. To account

[28]Most of my analyzes are also run at the state level, which eliminates any concern of time comparability. The results of the analysis that follow are very similar for states and MSAs. In future work, I plan to improve the time homogenization and also compare my results with those conducted at the level of commuting zones (Refer to section 4.1.1 for a definition of commuting zones).

for this, I run the following regression on the Census and ACS data to create a composition adjusted wage measure (at least based on observables):

$$w_{ijt} = \gamma_t + \Gamma_t X_{it} + \epsilon_{ijt}$$

where w_{ijt} is the log of hourly wages of worker i living in MSA j at time t. The workers characteristics are grouped in the variable that I call X_{it}. The X_{it} includes dummies for age (21-30,31-40,41-50,51-55), one dummy for gender, a US-born dummy (whether the worker was US-born or not), and a series of race dummies with being white the omitted group. In my controls I do not include the education status of the worker since I am going to compute the skill premium for college graduate versus less than college graduate workers.

Migration Rates I construct migration rates using data from March CPS. The reason why I take this data is that they are better suited than the Census data for this task. Unfortunately, information on migration is quite sparse in the Census. My estimation sample consists of all individuals between 16 and 55 years of age for which I have observations for the years from 1962 to 2009 available in the March CPS, with the exclusion of 1972-1975 and 1977-1979. I compute the migration rate in two ways. First, I use information collected in the CPS. I code someone as migrant if they migrated from a different MSA within the last year. I count all the workers that migrated by year, highly skilled (yes or no), and MSA weighted by their population shares in the MSA. Then, I divide this number by the population in the MSA. This procedure gives me the migration share for each MSA by education for each year in the sample available from CPS. To make sure that my approach is robust to other ways of computing the migration shares, I also calculate the number of workers living in a MSA minus the number of workers that were actually born in that MSA. The population in the MSA then divides everything. The results that I will show in the next section are robust to both approaches. In order to avoid potential biases because of the change in composition of the labor force (besides education), I control for sex, age, race, and citizenship when I run regression 3.

A.2 Tables

Table 1: Wage Convergence Rates

Panel A: Convergence Rate

	(1) 1940-1980	(2) 1980-2010	(3) 1980-2010-IV	(4) 1940-1980	(5) 1980-2010	(6) 1980-2010-IV
Log hourly wage, 1940	-0.0127***			-0.0159***		
	(-12.44)			(-20.54)		
Log hourly wage, 1980		-0.00122	0.00333		-0.00891**	-0.00164
		(-0.25)	(0.46)		(-2.91)	(-0.25)

Panel B: Convergence Rate by Skills

	1940-1980		1980-2010	
	No College	College	No College	College
Panel B.1				
Log hourly wage, 1940	-0.0123***	-0.0149***		
	(-14.32)	(-12.63)		
Log hourly wage, 1980			-0.0169***	0.000638
			(-9.70)	(0.30)
Panel B.2				
Log hourly wage, 1940	-0.0143***	-0.0216***		
	(-16.48)	(-21.30)		
Log hourly wage, 1980			-0.0200***	-0.00785***
			(-12.31)	(-3.87)
N	132	132	247	247

Note: This table reports the estimates of the β-convergence plotted in figures 1 and 2. In Panel A, I report the estimate of the β coefficient for the whole sample underlying figure 1. In column (1), there are β estimates for 1940-1980, and the observations are population-weighted. Column (2) has the same estimation but for 1980-2010. In columns (4) and (5), the estimations are not population-weighted. Column (3)-(6) have the population-weighted(unweighted) estimates for the IV regression where wages in 1970 are the instrument. In Panel B.1, I report the estimates of the β-convergence corresponding to figure 2. In column (1), I report the estimate for less skilled workers for 1940 and 1980; in column (2), for college graduates in the same time period. In columns (3) and (4), the estimates are once again for the two groups, but for the 1980-2010 period. In Panel B.2, I report the same estimates as in Panel B.1, but the observations are not population-weighted. All the standard errors are robust. T-stats are in parenthesis. The ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels respectively. The dependent variable in each regression is the annual average wage growth between the initial and final year reported at the top.

Table 2: Skill Premium by College Ratio of Cities over Time

	(1) Skill Premium		(2) Skill Premium	
College Ratio in 1940	-0.0631	(-0.43)	0.0775	(1.29)
College Ratio in 1950	-0.0475	(-0.51)	0.0199	(0.30)
College Ratio in 1970	-0.0505	(-0.39)	0.0132	(0.10)
College Ratio in 1980	-0.0824	(-1.08)	0.0308	(0.39)
College Ratio in 1990	-0.267***	(-3.85)	-0.138	(-1.50)
College Ratio in 2000	0.0621	(0.85)	0.186	(1.93)
College Ratio in 2010	0.217**	(2.99)	0.316***	(3.45)
Population	0.100***	(7.52)		
Time fixed effects	yes		yes	
N	1480		1480	

Note: This table reports the coefficients for the OLS. The dependent variable is the skill premium measured as the difference between the log wages of college graduates and less skilled workers. The only difference between column (1) and column (2) is that I control for population in column (1). The t-statistics are presented in parentheses. Observations are clustered at the state level. The ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels respectively.

Table 3: Migration over Time by College Ratio of Cities by Year

	(1) Migrant		(2) Migrant	
Migrant				
Coll. Ratio*High Skill in 1964	0.0275	(1.07)	0.0136	(0.51)
Coll. Ratio*High Skill in 1965	0.0744***	(4.63)	0.0589***	(3.54)
Coll. Ratio*High Skill in 1966	0.0590***	(3.45)	0.0481**	(3.02)
Coll. Ratio*High Skill in 1967	0.102***	(5.35)	0.0926***	(5.25)
Coll. Ratio*High Skill in 1968	0.0997***	(5.41)	0.0920***	(4.87)
Coll. Ratio*High Skill in 1969	0.0918***	(3.32)	0.0799**	(2.99)
Coll. Ratio*High Skill in 1970	0.0697***	(5.61)	0.0630***	(4.81)
Coll. Ratio*High Skill in 1971	0.0886***	(5.53)	0.0770***	(4.66)
Coll. Ratio*High Skill in 1976	0.0398	(1.38)	0.0238	(0.81)
Coll. Ratio*High Skill in 1980	0.221***	(3.90)	0.212***	(3.76)
Coll. Ratio*High Skill in 1981	0.0983***	(3.54)	0.0882**	(3.07)
Coll. Ratio*High Skill in 1982	0.134**	(3.27)	0.125**	(3.00)
Coll. Ratio*High Skill in 1983	0.0779***	(5.35)	0.0728***	(4.83)
Coll. Ratio*High Skill in 1984	0.0951***	(6.03)	0.0898***	(5.10)
Coll. Ratio*High Skill in 1985	0.193***	(3.37)	0.193***	(3.31)
Coll. Ratio*High Skill in 1986	0.0897***	(6.06)	0.0854***	(5.73)
Coll. Ratio*High Skill in 1987	0.0708**	(2.85)	0.0719**	(2.96)
Coll. Ratio*High Skill in 1988	0.0688***	(3.52)	0.0693***	(3.62)
Coll. Ratio*High Skill in 1989	0.0791***	(4.23)	0.0798***	(4.29)
Coll. Ratio*High Skill in 1990	0.0795***	(4.94)	0.0813***	(5.16)
Coll. Ratio*High Skill in 1991	0.0601**	(2.70)	0.0644**	(2.82)
Coll. Ratio*High Skill in 1992	0.118***	(4.86)	0.105***	(4.33)
Coll. Ratio*High Skill in 1993	0.107***	(4.02)	0.0942***	(3.53)
Coll. Ratio*High Skill in 1994	0.115***	(5.29)	0.108***	(4.89)
Coll. Ratio*High Skill in 1995	0.0136	(0.54)	0.00593	(0.23)
Coll. Ratio*High Skill in 1996	0.123***	(6.07)	0.108***	(5.22)
Coll. Ratio*High Skill in 1997	0.0971***	(4.63)	0.0857***	(4.02)
Coll. Ratio*High Skill in 1998	0.133***	(6.66)	0.120***	(5.77)
Coll. Ratio*High Skill in 1999	0.103***	(4.69)	0.0939***	(4.21)
Coll. Ratio*High Skill in 2000	0.122***	(3.40)	0.112**	(2.97)
Coll. Ratio*High Skill in 2001	0.0817**	(2.87)	0.0757**	(2.60)
Coll. Ratio*High Skill in 2002	0.124***	(4.62)	0.116***	(4.35)
Coll. Ratio*High Skill in 2003	0.0828**	(2.62)	0.0771*	(2.38)
Coll. Ratio*High Skill in 2004	0.0927***	(3.39)	0.0863**	(3.02)
Coll. Ratio*High Skill in 2005	0.0792**	(3.22)	0.0714**	(2.87)
Coll. Ratio*High Skill in 2006	0.0974***	(3.98)	0.0915***	(3.70)
Coll. Ratio*High Skill in 2007	0.0986***	(4.23)	0.0928***	(3.95)
Coll. Ratio*High Skill in 2008	0.115***	(5.28)	0.108***	(4.87)
Time fixed effects	yes		yes	
Controls	No		yes	
N	1411802		1411802	

This table reports the marginal effects for every year for the probit regressions. The dependent variable is the decision on whether to move or not. Standard errors are presented in parentheses and are clustered at the state-level. The ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels respectively. Column (2) is identical to column (1) except that column (1) controls for population.

Table 4: $\Delta \frac{H}{L}$ vs. Initial $\frac{H}{L}$ in the Data

	(1) 1940-1950	(2) 1950-1970	(3) 1970-1980	(4) 1980-1990	(5) 1990-2000	(6) 2000-2010
Panel A						
$\frac{H}{L}$	-0.218*	-0.439***	0.0355	-0.00158	0.0708***	0.0401*
	(0.115)	(0.0887)	(0.0587)	(0.0305)	(0.0238)	(0.0218)

	1950-1970	1950-1980	1970-1990	1980-2000	1990-2010
Panel B					
$\frac{H}{L}$	0.240**	-0.320***	0.0970	0.0770**	0.0797**
	(0.117)	(0.0963)	(0.0808)	(0.0390)	(0.0386)
N	103	143	119	247	238

Note: Panel A shows the estimates of running the initial $\frac{H}{L}$ on the growth over 10 years, $\Delta \frac{H}{L}$. Panel B replicates the same analysis as Panel A for the growth over 20 years, $\Delta \frac{H}{L}$. Standard errors are in brackets. The ***, **, and * represent statistical significance at the 0.001, 0.01, and 0.05 levels respectively.

Table 5: First-Stage Estimates of Models for Routine Occupation Share Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dep. Variable S_{Hjt}						
$\Delta \hat{S}_{Hjt}$	3.046***	3.643***	2.852***	4.418***	3.062***	3.043***
	(0.620)	(1.024)	(0.632)	(1.118)	(0.719)	(0.737)
F	24.12	12.65	20.34	15.63	18.14	17.06
Panel B: Dep. Variable S_{Ljt}						
$\Delta \hat{S}_{Ljt}$	1.021***	0.891**	0.850***	2.483***	2.535***	2.511***
	(0.341)	(0.344)	(0.285)	(0.531)	(0.527)	(0.591)
F	8.975	6.709	8.891	21.86	23.15	18.06
N	144	119	270	249	283	283

Note: In this table I report the first-stage estimates between the instrumental variable and the measure of skill bias. Standard errors are in brackets. In column (1), I run the regression for 1950 and in column (2)-(6) for 1970-2010. The sample does not contain 1960. In panel A, I report the results for highly skilled workers and in panel B for less skilled workers. The ***, **, and * represent statistical significance at the 0.001, 0.005, and 0.01 levels respectively.

Table 6: Model Estimates for 1940-2010

		(1)	(2)	(3)	(4)	(5)
A. Labor Supply						
θ	Share of Housing	0.618***	0.608***	0.431***	0.446***	0.421***
		(0.102)	(0.100)	(0.117)	(0.112)	(0.116)
γ^p	Elast. w.r. t. to $\frac{H}{L}$			0.914***	0.940***	0.927***
				(0.119)	(0.113)	(0.122)
B. Labor Demand						
ρ	$\frac{1}{1-\rho}$=elast. of substitution	0.603***	0.748***	0.526***	0.319*	0.670***
		(0.226)	(0.183)	(0.165)	(0.166)	(0.194)
γ^H	Wage Elast. w.r.t. $\frac{H}{L}$ for H	0.722***	0.545*	0.723***		0.528**
		(0.280)	(0.280)	(0.252)		(0.259)
γ^L	Wage Elast. w.r.t. $\frac{H}{L}$ for L	0.073	0.056	0.118		0.113
		(0.245)	(0.262)	(0.217)		(0.223)
ϕ^H	Wage Elast. w.r.t. $(H+L)$ for H		0.271***			0.253***
			(0.080)			(0.079)
ϕ^L	Wage Elast. w.r.t. $(H+L)$ for L		0.210			0.211
			(0.133)			(0.141)
λ^H	Wage Elast. w.r.t S^H for H	-0.096	-0.048	-0.083	0.072	-0.067
		(0.109)	(0.128)	(0.105)	(0.123)	(0.113)
λ^L	Wage Elast. w.r.t S^H for L	-0.237**	-0.160	-0.228*	-0.998***	-0.160
		(0.116)	(0.143)	(0.128)	(0.236)	(0.150)
A. Diffusion						
γ^2	Tech. Diffusion	0.998***	0.934***	0.997***	1.190***	0.951***
		(0.051)	(0.067)	(0.051)	(0.035)	(0.064)
Observations		705	705	705	705	705

Note: In this table, I report the moments and the estimates of the model. The ***, **, and * represent statistical significance at the 0.001, 0.005, and 0.01 levels respectively.

Table 7: Externally calibrated Parameters

Parameter	Value	Literature
Subsistence level of Housing: $\bar{O}$	0.25	Ganong and Shoag (2017)
Elasticity of Supply Housing: μ	0.4	Ganong and Shoag (2017)
Migration costs: σ^L and β^L	-.065 and -.861	Notowidigdo (2011)
Migration costs: σ^H and β^H	-.066 and -1.044	Notowidigdo (2011)
Share of each Intermediate: α	0.51	Burstein et al. (2017)

Table 8: $\Delta\frac{H}{L}$ vs. Initial $\frac{H}{L}$ in the Model

	(1) 1940-1950	(2) 1950-1960	(3) 1960-1970	(4) 1970-1980	(5) 1980-1990	(6) 1990-2000	(7) 2000-2010
$\frac{H}{L}$	-0.245***	-0.244***	-0.244***	-0.212***	0.332***	0.170***	0.0826***
	(0.00248)	(0.00259)	(0.00271)	(0.00861)	(0.0289)	(0.00983)	(0.00493)

Note: Column (1) shows the estimates of running the initial $\frac{H}{L}$ in 1940 on the growth over the 30 years, $\Delta\frac{H}{L}$ between 1940 and 1970. Columns from (2) to (7) show the estimates of running the initial $\frac{H}{L}$ on the growth over 20 years for each period from 1960-1980 until 1990-2010.

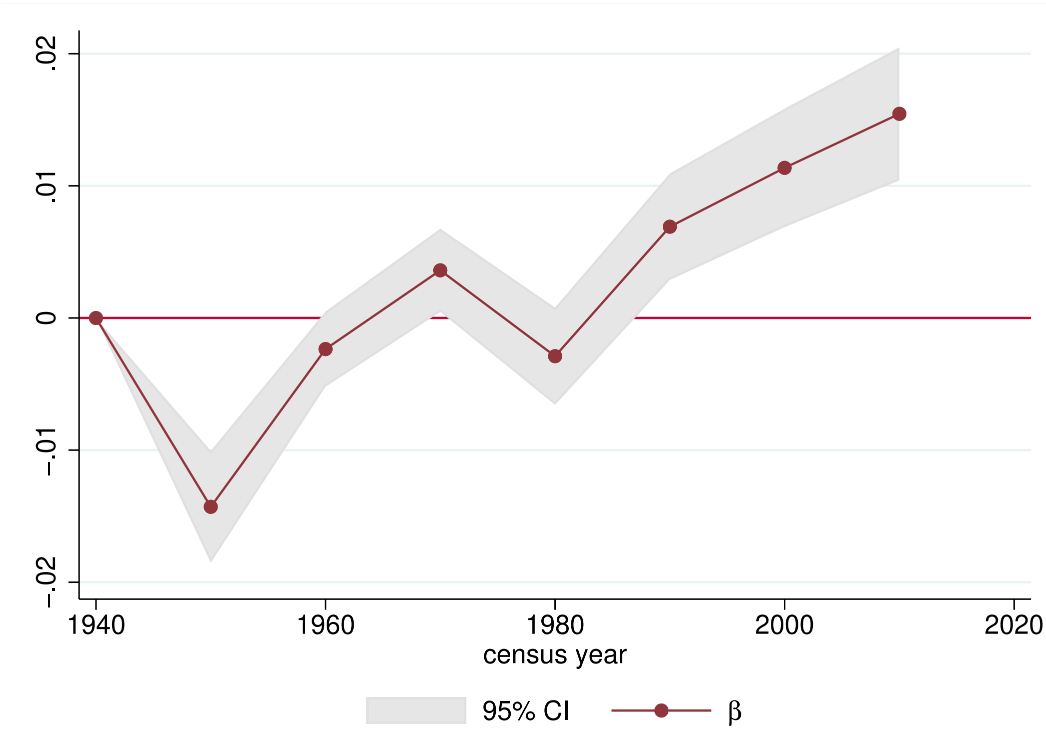
Table 9: Change in β, β^H and β^L estimates in the Model over Time

	Full	**No Agglom.**	**No SBTC**	**No Housing**	**No Migr. Cost**
Panel A: Singular Decomposition					
$\Delta\beta$	-100.37	-17.63	-34.28	-96.16	-97.50
$\Delta\beta^H$	-187.67	-76.63	-62.55	-183.03	-185.51
$\Delta\beta^L$	-23.52	35.55	-8.41	-20.30	-19.99
Panel B: Sequential Decomposition					
$\Delta\beta$	-100.37	-17.63	-47.13	-47.73	-46.35
$\Delta\beta^H$	-187.67	-76.63	-74.87	-76.68	-73.77
$\Delta\beta^L$	-23.52	35.55	-21.91	-21.64	-21.58

Note: Both in Panel A and B, the first row of the table shows the results for the change in β overall between 1980 and 2010 in the model. The second(third) row has the results for the increase in β^H (β^L) in the model. In panel A, I remove one at the time the different channels from the model. In panel B, in each column I vary from the full model to removing step-wise the elements of the model.

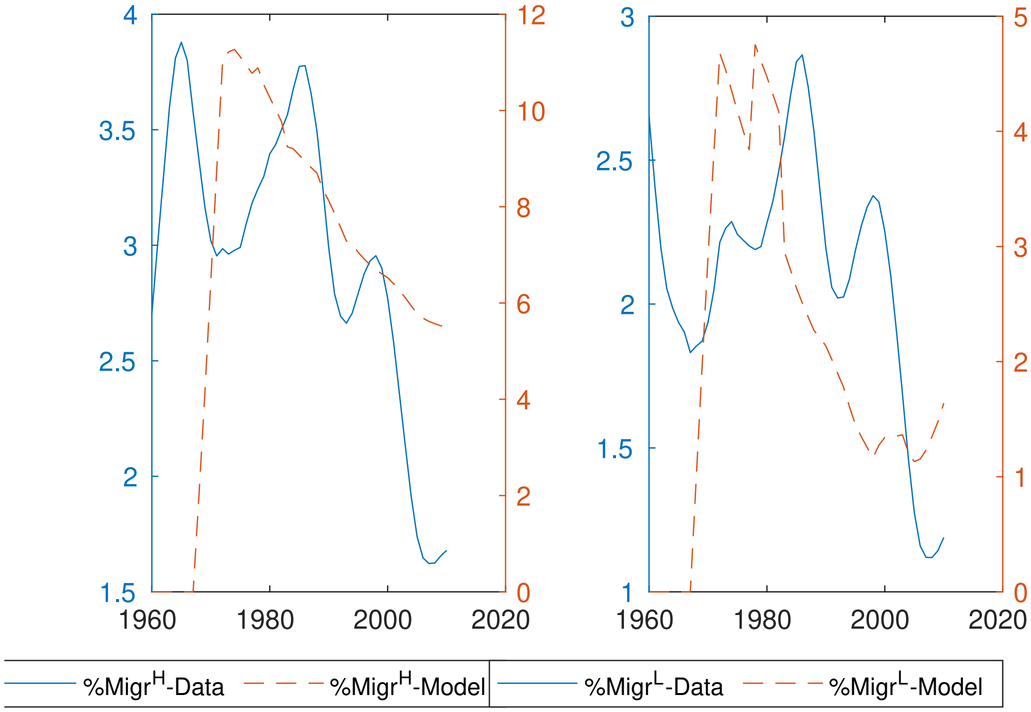
A.3 Figures

Figure 12: Skill Premium by MSA Population Levels



Note: This figure plots the estimate of the coefficient β for the regression 2. On the horizontal axis, I have the decades from 1940 to 2010, while on the vertical axis, I have the estimate of coefficient β for each decade from 1940 to 2010. Moreover, there is a line starting at zero on the vertical axis.

Figure 13: Migration Rate Over Time



Note: This figure shows the evolution of the migration rate for highly skilled and less skilled workers over time for both the model and the data. On the left, I plot the migration rates generated by the model with a cross and those generated by the data with a circle. On the right plot, I plot the migration rates for less skilled workers.

Figure 14: Time Evolution of Rents and Prices Variance

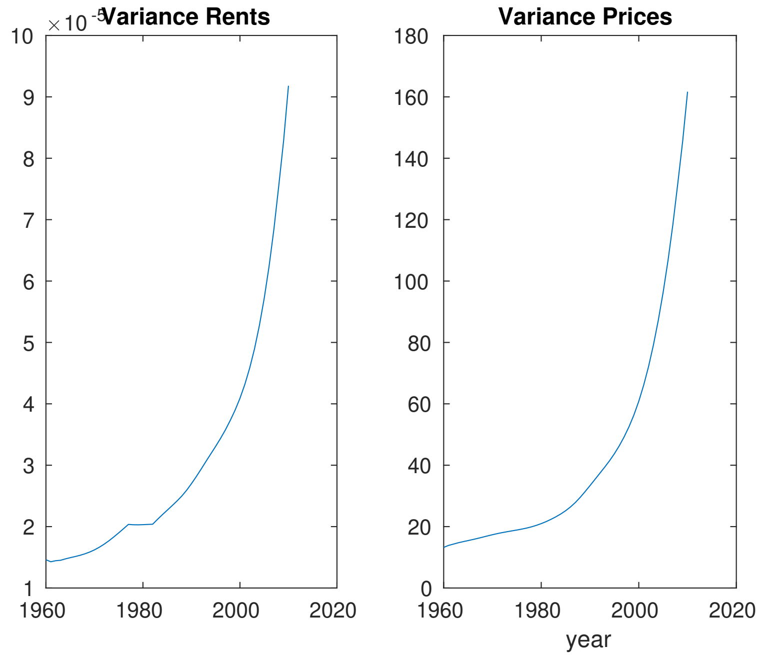
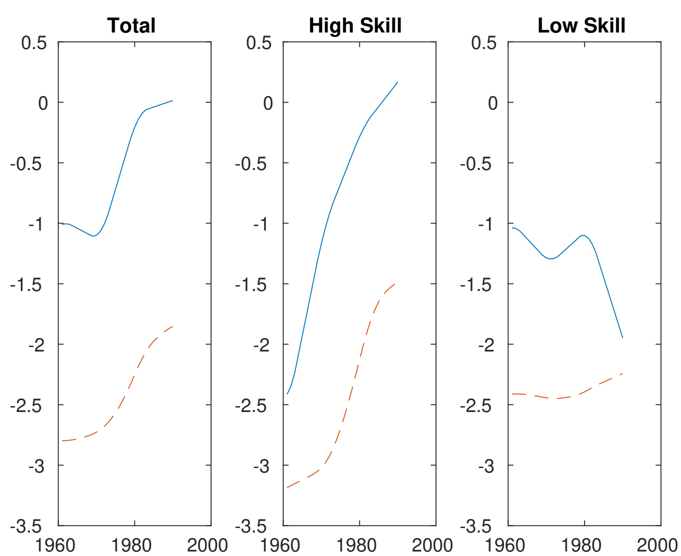
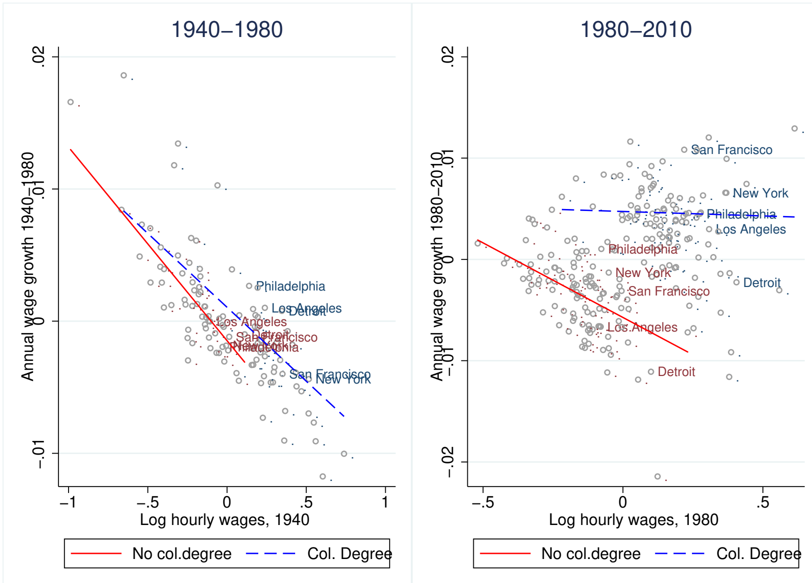


Figure 15: Model Matching Data on Real Wage Convergence



Note: This figure on the left (center)(right) shows a rolling estimate of the $\beta(\beta^H)(\beta^L)$-convergence with a rolling window of 20 years. The solid line is the data for which I smooth the 20-year rolling estimate, while the dashed line is the estimate of the β-convergence from the model for which I compute a yearly estimate.

Figure 16: Wage Convergence across MSAs before and after 1980 by Skill Group - Low housing elasticity states



Note: This figure shows two scatter plots of log wages by MSA in the initial year against the annual average growth of wages in the final year by skill type (highly skilled and less skilled workers) in cities that are in states with low housing elasticities. Specifically, on the left-hand side (right-hand side), I plot the demeaned log wages in 1940 (1980) by MSA against the annual average growth of wages between 1940 (1980) and 1980 (2010) by skill type (highly skilled and less skilled workers). The line in each graph represents a weighted regression line from the bi-variate regression.

A.4 Equilibrium and Discussion of the Assumptions

I define the dynamic competitive equilibrium of this model as follows:

Definition The equilibrium consists of a set of allocations $\{\{L_{djt},\, H_{djt}\}_{d=1}^{D}\}_{j=1}^{J}$ and a set of prices $\{\{P_{djt}\}_{d=1}^{D},\, R_{jt}\}_{j=1}^{J}$, wages $\{W_{Hjt}, W_{Ljt}\}_{j=1}^{J}$, such that given $\{\{\xi_{Ldj0},\, \xi_{Hdj0}\}_{d=1}^{D}\}_{j=1}^{J}$, $\{\{A_{Ljt},\, A_{Hjt}\}_{t=1}^{T}\}_{j=1}^{J}$, a set of parameters normalizing $P_{jt} = P_t = 1$ and $\sum_j (L_{jt} + H_{jt}) = 1$ in each time period t:

1. Given migration costs and idiosyncratic preferences, workers choose their location and consumption to maximize the utility satisfying equations 7, 8 and 9;

2. Firms maximize profits such that equations 12, 13 hold;

3. There is free entry for firms into the tradable sector such that $\pi = 0$;

4. Labor markets clear such that 7 and 8 hold;

5. Housing markets clear such that the demand of equation 15 is equal to the supply in equation 14:
$$R^{\mu}_{jt} = H_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Ljt}}{R_{jt}}\right]$$
6. Total labor supplies are the sum of labor demanded in each intermediate such that
$$L_{jt} = \sum_{d=1}^{D} L_{djt} \qquad \text{and} \qquad H_{jt} = \sum_{d=1}^{D} H_{djt}$$
7. Technology evolves according to 11.

A.5 Existence and Uniqueness

Because of the endogenous productivity channels, this model might allow for multiple equilibria. These equilibria might happen if the agglomeration forces are strong enough that the workers agglomerate all together in the same locations. To avoid this problem, I must impose restrictions on the parameters governing the production function such that the agglomeration forces are compensated for by dispersion forces. Allen and Arkolakis (2014) prove the existence and uniqueness of an equilibrium in a static model with agglomeration forces. Desmet et al. (2018) extend the proof to a dynamic model with only one type of agent. Both studies find that the strength of agglomeration and dispersion externalities are crucial in guaranteeing the uniqueness and existence of a spatial equilibrium. Unfortunately, the proofs of Allen and Arkolakis (2014) and Desmet et al. (2018) do not apply and cannot be extended to a case with heterogeneous labor aggregated in a CES fashion. Therefore, I proceed with solving the model for several sets of agglomeration parameters. These simulations show that the values of the agglomeration parameters for which the model has multiple equilibria are definitely higher than the ones I estimate in section 4.1.

A.6 Discussion

Spatial Technology diffusion: Introducing this persistent productivity formulation with spatial diffusion generates convergence directly in the model, as in Barro and Sala-i Martin (1997), Caselli and Coleman (2001), and Desmet et al. (2018). Unlike a model that compares steady-states, convergence generated with a diffusion mechanism is better suited to the explanation in Barro and Sala-i Martin (1997) that argues that a neoclassical model with friction to capital mobility reproduces the convergence rates across countries and within the US. Caselli and Coleman (2001), instead, build a dynamic model in which total factor productivity (TFP) grows faster in agriculture, there are declining costs of acquiring human capital, and farm goods are a necessary good. These two models introduce convergence through two different mechanisms. Also, Caliendo (2011) and Bajona and Kehoe (2010) show that convergence can be proven in a dynamic Hecksher-Ohlin model. The convergence produced by an idea-diffusion process is closer to a declining cost of human capital or to physical capital mobility. As formerly mentioned and shown in appendix E, the model without spatial technology diffusion is not able to replicate the main results.

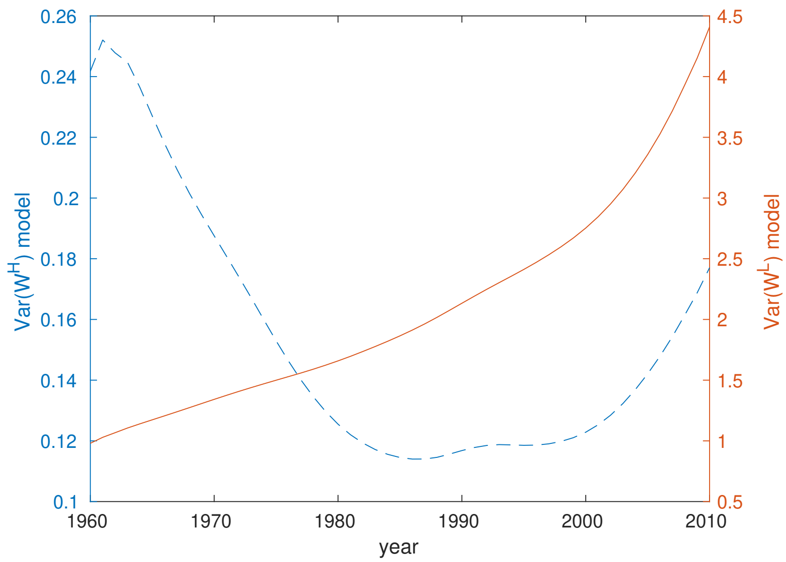
SBTC: I do model SBTC as an exogenous shock to productivity that differs between the two skill groups. Katz and Murphy (1992) in their seminal work think of skill-biased technical change as a residual in productivity that changed over time. In this paper, I use the routinization measures, similar to the ones developed in Autor and Dorn (2013) to distinguish technological innovation from other residuals. Autor and Dorn (2013) use similar measures to capture polarization instead. The literature on SBTC has used several approaches to model it. Krusell et al. (2000) and Beaudry et al. (2010) model it as a capital-labor relationship. In Appendix D.2, I show that a model with physical capital and decrease in the price of computers would not reproduce some features that are present in the current setting and that are key in the data. On top of that, the current setting incorporates industrial composition while not accounting for physical capital. While physical capital is important in the production of goods, it is not crucial for the purposes of this paper. But, how would physical capital bias the results of this model? This answer depends on the mobility of capital and on the complementarity or substitutability of capital with highly skilled labor. If physical capital is freely tradable such that rental rates are equalized across locations, then the model would draw the same conclusions as it does without capital.

Parameters' set: On the production side, the spillover's parameters could be more parsimonious and reduced to three rather than four. However, the current formulation allows to have a closer connection to existing estimates and determine the elasticities for both groups of workers and both for the skill ratio and the size effect. Qualitatively, no key moment would be affected. In the housing market there are two parameters. These serve to match the Stone-Geary preferences as in Ganong and Shoag (2017). Even in the absence of the housing mechanism, the model would reproduce key moments of the data as shown later.

B Other Results of the Model

In this section, I report results comparing model and data on trends in wage dispersion and in the "Great Divergence" of skills.

Figure 17: Variance of High and Less Skill Model over Time



B.1 Wage Dispersion Increase Over Time

Hsieh and Moretti (2015) show that wage dispersion across US cities increased substantially between 1964 and 2009. As in table 10, the model shows that wage dispersion in the US has increased substantially over the last 30 years in accordance with the findings in Hsieh and Moretti (2015). My model supplements this finding by predicting differences in wage dispersion between highly skilled and less skilled workers. Figure 17 suggests that the variance in the highly skilled wages was decreasing until 1980 when it started increasing. Instead, the variance of wages for the less skilled group was constantly increasing over time. As shown in table 10, the variance between 1964 and 2009 increased by 213%. It only increased 5% for the highly skilled while it increased by 222% for the less skilled. I run some counterfactual analysis by shutting down agglomeration forces as in section 5.2 sequentially and I find that if the agglomeration economies had been set to 0, then, the increase in wage variance would have gone up only by 11.43% for the whole group, but it would have decreased for highly skilled as shown in column (2) and mildly increased for the low-skilled workers. In column (3), I also shut down the SBTC finding that variance would have decreased both for the highly and for less skilled workers if sone of these forces had been in place. Column (4) and (5), respectively, housing costs and migration, suggest that the contribution to variance in wages is close to null, quantitatively.

Table 10: Change in Variance of High and Less Skill Model over Time in the Model

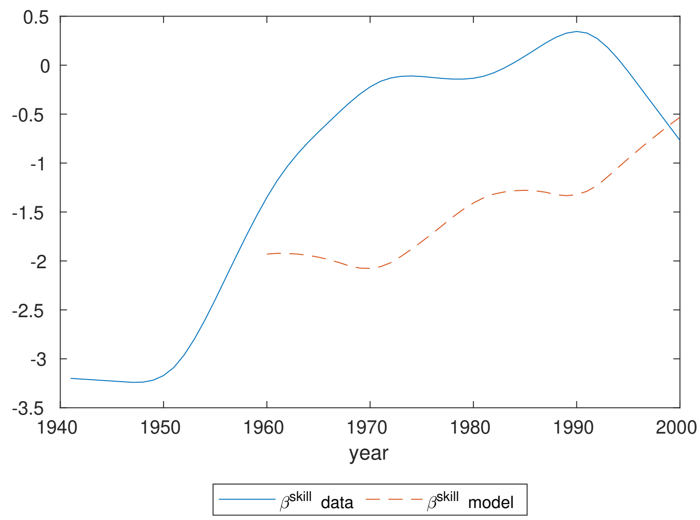
	Full	**No Agglom.**	**No SBTC**	**No Housing**	**No Migr. Cost**
$\Delta Var(W)$	213.21	11.43	-59.50	-59.41	-59.44
$\Delta Var(W_H)$	5.10	-63.41	-47.91	-47.29	-47.88
$\Delta Var(W_L)$	222.84	21.46	-61.21	-61.19	-61.16

Note: The first row of the table has the results for the increase in wage dispersion overall between 1964 and 2009 in the model. The second(third) row has the results for the increase in wage variance for highly and lees skilled workers between 1964 and 2009 in the model. In each column I vary from the full model to removing step-wise the elements of the model.

B.2 The "Great Convergence and the "Great Divergence" of Skills over Time

Using model-generated data, I estimate specification 4 and compare the outcomes in the model and in the data in figure 18. The model reproduces shows a β^{skill} convergence rate going from -2.5% in 1970 to -.5% in 2010. Overall, β^{skill} declines much faster despite the last period, where it coincides with the model. Overall, this finding suggest that the model is consistent with this other non-targeted moment, which means that it reproduces not only features of the wage data, such as the decline in cross-MSA wage convergence, but also features of employment data, such as the divergence in the skill ratio. In table 8, when I compare the 10-year rolling window estimates over time, the coefficients slowly moves from -.245 in 1940-1950 to 0.0826 in 2000-2010, switching sign in 1980-1990.

Figure 18: The "Great Divergence" in Skills: Data vs. Model



Note: This figure shows the evolution of the estimates of β_t^{skill} in equation 4 both in the data and in the model.

C Other Potential Explanations

There are several potential explanations that are complementary to the SBTC and agglomeration conjecture. In this section, I explore the changes in policy such as housing, unionization, *Right to Work Laws*, and international trade together with the industry's composition and the firms' decisions on location.

C.1 Housing Regulation

Ganong and Shoag (2017) provide a explanation based on housing prices that suggests that the US states where housing prices increased the most are also the ones where the migration declined. Hence, because migration increases convergence, the decline in migration to this area, which is also the richest, also decreased the income convergence rate. As suggested in their paper, the housing prices and SBTC could be complementary. For this reason, in order to decide how to disentangle them, I add a housing sector to the model to compare the housing effects with my key mechanisms.

Additionally, I conduct an empirical test that shows that even in the areas where the housing restrictions are high, there is a strong difference in the convergence rate of wages for the highly skilled and the less skilled groups. I construct figure 2 for the MSAs that were in states where the housing prices went up dramatically because of housing regulations. Figure 16 shows that the effect of regulations on the decline in income convergence looks quite similar to the one without any restriction. Thus, I can conclude that there is room also for the SBTC in the group of states where housing prices are high.

C.2 Innovation and Financial Sector

Another potential and complementary explanation is that technological innovation might have caused a sectoral effect rather than a skill-biased effect. Such an effect would cause productivity increases in highly innovative industries such as communication. Therefore, cities with a higher concentration of innovative industries benefit more from the technological change. To control for the importance of the IT sector, I estimate conditional convergence in wages between 1980 and 2010.[29] The results reported in Table 11 show that unconditional wage convergence is not statistically significant in column A. However, when I add a control for the IT sector in column B, the coefficient for wages in 1980 becomes positive and statistically significant. In column C, I add a control for highly skilled wages, and the coefficient on initial wages in 1980 increases in magnitude. This evidence shows that adding sectoral differences in technological intensity have the effect of amplifying the decline in spatial convergence. The framework developed above takes into account these sectoral differences by including a highly skilled and less skilled sectors.

In addition to sectoral innovation shifts, changes in firms' relocation decisions over time can contribute to the decline in wage convergence. More skilled firms might begin to move to richer places but then reverse their decision and move to poorer cities to take advantage of lower costs. In order to investigate whether firms' location decisions change over time requires firm-level data. Faberman and Freedman (2016) use longitudinal establishment data for the US during the years 1992-1997. They do not find that spillover is important for firms' decisions to locate in urban areas rather than other areas. Unfortunately, the data on the firms' locations back to 1940 are not available. In this regard, I use publicly available data at the industry level to test whether more skilled occupations have become increasingly concentrated in more skilled cities over time. If this is the case, it might mean that in addition to the sorting of highly skilled workers into highly skilled cities, there is also sorting of highly skilled firms into highly skilled cities. To test this hypothesis

[29] I define IT sector by looking at the codes of the IND1990 variable in the IPUMS data set and select industries that are more technology-oriented.

empirically, I run the following regression to obtain the marginal effects by decade

$$\text{Skill concentration}_{kjt} = \alpha + \sum_{t=1950}^{T} \beta_t \left(\frac{H_{jt}}{L_{jt}}\right) + \phi_t + \phi_j + \epsilon_{kjt}$$

where k is the industry, j is the MSA, and t is time. The ϕ_t are time fixed effects, and ϕ_j are MSA fixed effects. I build the measure of "Skill concentration" by calculating the ratio between the number of skilled workers over the number of total workers that are in industry k in location j at time t. This hypothesis is confirmed in the data. In figure 19, I plot the coefficient β_t over time. The figure shows that a more skill-concentrated MSA becomes more strongly correlated with skill concentration at the industry level. This concentration is evidence of sorting not just of workers but also of industries and thus, firms.

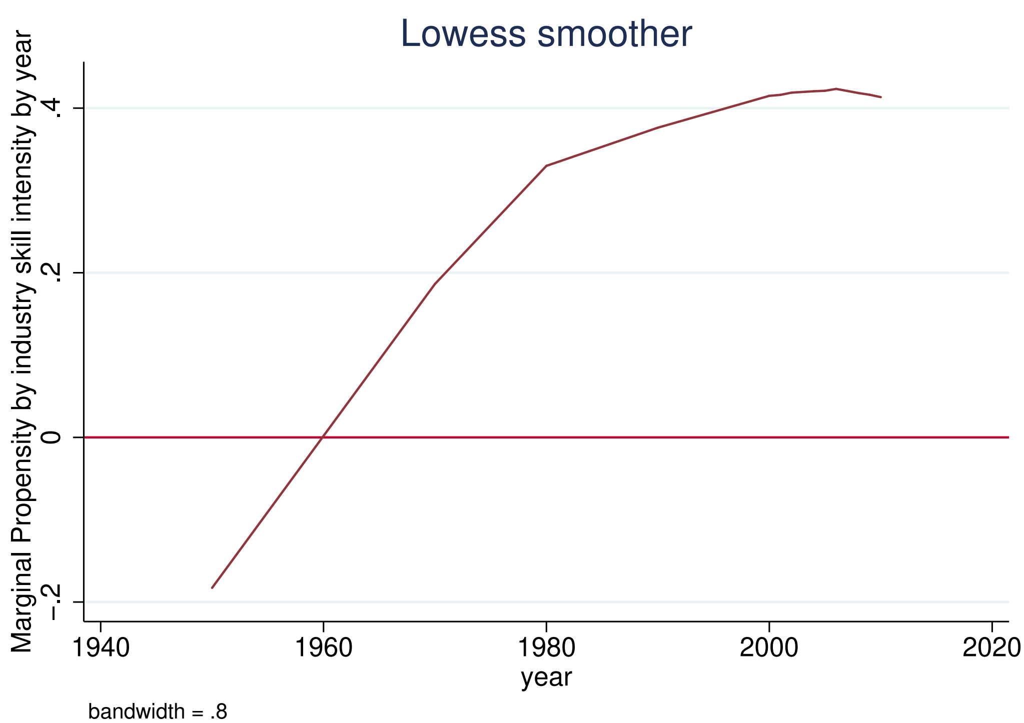
C.3 Right to Work Laws and Unions

In Southern and Western US, 26 states have passed *Right to Work Laws* since 1940. These laws permit workers to work without having to join a union. The *Right to Work Laws* might have a spatial effect of increasing the wages of less skilled workers in the states where they were implemented. In fact, Holmes (1998) shows that state policies play a role in the location of an industry. However, only 26 states have adopted right to work laws and figure 20 shows that the majority of the states passed these laws in the 1950s and 1960s, long before the secular decline in wage convergence. Besides the *Right to Work Laws*, union membership has gone up substantially in the US, and this growth might have directly affected the wage convergence rate. In order to account for this growth, I use data on unions from CPS survey aggregated at state level starting in 1990. Table 12 reports the estimates of wage convergence's regression at the state level between 1990 and 2010. I find that controlling for the presence of unions does not increase the β estimates. If anything, it actually decreases it.

C.4 International Trade

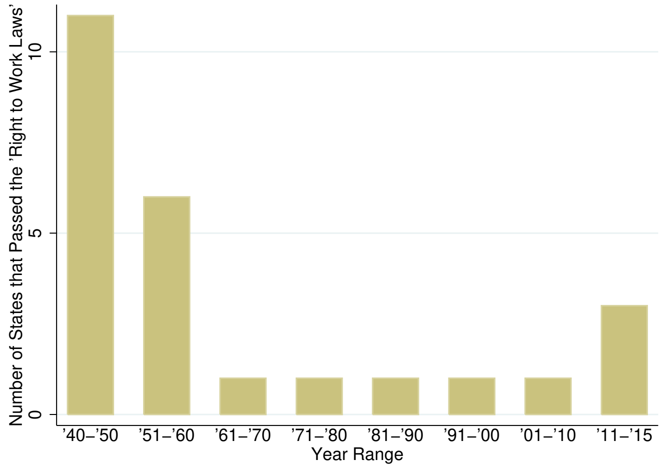
Besides SBTC, there is evidence that international trade has an effect on the increase in the skill premium at national level (Feenstra and Hanson 1999). As a consequence, this trade might have an effect on the slowdown in regional convergence as well. However, Feenstra and Hanson (1999) find that there is no large effect of international trade, such as outsourcing, on the skill premium. Instead, Autor et al. (2015) find that the increase in the import penetration from China affected employment rates in the commuting zones where the penetration from China was higher. However, outsourcing had modest effects on the skill premium and imports from China. Therefore, while this demand might also be relevant, the timeline and the magnitude of the effect does not explain the strong slowdown in the regional convergence of wages. However, I do run the wage convergence's regression at the state level while controlling for the import penetration from China as in Autor et al. (2015). The results show that controlling for the trade shock does not affect the speed at which wages converge.

Figure 19: Industry Sorting over time



Note: This figure plots the estimated effect of skill concentration at the MSA level and at the industry level. The line is computed using the estimates of the skill ratio at the MSA level (β), using specification C.2.

Figure 20: Right To Work Laws



Note: This histogram plots the number of states that passed the "Right to Work Laws" by decade starting with the 1940s.

Table 11: Convergence Rates by Skills and IT

	(1) A	(2) B	(3) C
Log hourly wages 1980	-0.0000389	0.00593**	-0.0126***
	(-0.02)	(2.95)	(-10.58)
IT		0.00656***	0.00538***
		(13.49)	(16.54)
col_degree			0.0106***
			(19.85)

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The dependent variable in this table is Δw_{jt} for location j at time t. The initial period is 1980 and the final period is 2010. In column A, I run it against wages in the initial period 1980. In column B, I control for the IT sector dummy. In column C, I control for the college degree.

Table 12: Wage Convergence with Unions Controls

	(1)	(2)	(3)	(4)
Log hourly wage, 1980	-0.00840***	-0.00723**	-0.00604**	-0.00454
	(-2.76)	(-2.31)	(-2.28)	(-1.63)
Union		-0.0413**		-0.0427**
		(-2.27)		(-2.28)
Pop. Weight	Yes	Yes		
Observation	147.00	147.00	147.00	147.00
R square	0.26	0.30	0.14	0.19

Note: This regression shows the coefficient for the decline in wage growth between 1990 and 2010 on the initial wage in 1990, conditioning on the union presence by state. All the observations are clustered at the state level.

Table 13: Wage Convergence with Import Penetration from China

	(1)	(2)	(3)	(4)
	(1)	(2)	(3)	(4)
Log hourly wage, 1980	-0.00840***	-0.00760**	-0.00604**	-0.00539**
	(-2.74)	(-2.58)	(-2.26)	(-2.12)
China Import Penetr.		-0.000320		-0.000373
		(-1.34)		(-1.42)
Pop. Weight	Yes	Yes		
Observation	49.00	48.00	49.00	48.00
R square	0.26	0.27	0.14	0.16

Note: This regression shows the coefficient for a regression of wage growth between 1990 and 2010 on the initial wage in 1990, conditioning on the import penetration from China by state. All the observations are clustered at the state level.

D Theory Appendix

This appendix supplements the theoretical framework presented in Section 3 in several respects. In subsection D.1, I describe the algorithm for solving the system of equations and obtaining the solution to the model. Subsection D.2 presents a version of the model in which skill-biased technology, instead of being a local exogenous shock, is modeled as endogenous technology adoption. And, subsection D.3 derives an alternative expression for Y_T.

D.1 Description of the Computational Algorithm

In order to recover the equilibrium quantities and prices for period t, it is necessary to solve the full model numerically. I can reduce the equilibrium conditions by the following six, which are reported again below for the sake of clarity:

$$W_{Hjt} = (\eta_{Hdjt})[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}H^{\rho-1}_{djt} \tag{20}$$

$$W_{Ljt} = (\eta_{Ldjt})[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}L^{\rho-1}_{djt} \tag{21}$$

$$R^{\mu}_{jt} = H_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{22}$$

All $d \in D$ intermediate market sectors clear:

$$\frac{P_{djt}}{\alpha P_{jt}} = [\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}}$$

From the decision on the location of the labor market , labor market clearing becomes

$$H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))} \tag{23}$$

$$L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{24}$$

where

$$\begin{aligned} \delta_{kjt} = \Big[&\theta \ln(W_{kjt} - R_{jt}\bar{H}) + \\ &(1-\theta)[\ln((1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + (1-\theta)\ln((1-\theta)(W_{kjt} - R_{jt}\bar{O})] + \\ &+ A_{kjt} + \gamma^{p}\ln(H_{jt}/L_{jt})\Big] \end{aligned} \tag{25}$$

and

$$L_{jt} = \sum_{d=1}^{D} L_{djt} \qquad \textit{and} \qquad H_{jt} = \sum_{d=1}^{D} H_{djt}$$

I end up with a system of 46 equations in 46 unknowns $\{W_{Hjt},\ W_{Ljt},\ H_{djt},\ L_{djt},\ R_{jt}\ P_{djt}\,\forall j \quad \text{and} \quad \forall d\}$ for each MSA. Since the analysis includes 240 cities and 14 industries, I have a system of 46x240=11,040 equations. I solve this system using an iteration algorithm. The algorithm consists of the following steps:

1. Given the set of parameters $\{\gamma^H, \gamma^L, \phi^H, \phi^L, \rho, \gamma^2, \lambda^H, \lambda^L, \theta, \gamma^p\}$, the sequences of S_t^H and S_t^L and the sequences of A_{Hjt} and A_{Ljt}, the initial productivity ξ_{Ldj0} and ξ_{Hdj0} for all j cities and for all industries d;
2. Start by guessing an allocation of $\{H_{dj0}, L_{dj0}\}_{j=1,d=1}^{J,D}$;
3. For each location, compute an equilibrium allocation h_j, output Y_{dj} wages W_{Hj} and W_{Lj} and P_{dj};
4. Using the information on prices, compute $\{H_{j,L_j\}_{j=1}^J}$
5. Check whether the distance between the values of $\{H_j, L_j\}_{j=1}^J$ and the guesses $\{H_{j0}, L_{j0}\}_{j=1}^J$ are smaller than an exogenously given tolerance level equal to e^{-10}.
6. If so, then stop. If not, consider $\{H_j, L_j\}_{j=1}^J$ as the new guess and restart the loop. Continue the procedure until the distance is smaller than the tolerance level e^{-10}.

I solve the model for 70 periods where time t is a year. In the first 40 periods, S_{Ht} and S_{Lt} are set to zero, then I set the value of S_{Ht} and S_{Lt} from the data (at national level) from S and λ from the model estimation. Start looking for the equilibrium at time $t = 0$ and give a value for ξ_{j0}^H and ξ_{j0}^L where $\xi_{j0}^H > \xi_{j0}^L$ for all j generated by the estimation of the residuals of the wage equations in year 1940.

Although the complex structure of the model does not allow me to derive conditions under which the algorithm converges to an equilibrium distribution of population, simulation results indicate that the algorithm displays good convergence properties unless either agglomeration or dispersion forces are very strong. Specifically, the algorithm always converges to equilibrium in a broad neighborhood around the parameter values chosen in the calibration.

D.2 Model with Endogenous Innovation Rate

The model specified above provides for a SBTC that is exogenous and differs for each location j. However, I could allow SBTC to be modeled as "technological adoption" following Beaudry et al.

(2010). When computers arrive, firms need to decide whether to adopt them (PC) or stick with their current technology (K). This new technology is assumed to be skill-biased relative to the old technology because for the same level of prices, the new technology uses skilled labor more intensively. Specifically, where there is a higher concentration of highly skilled workers, there is also a higher ratio of computers per worker.

The production function with the old technology K is equal to

$$Y_d = \mathrm{K}^{(1-\alpha)}[aH^{\rho} + (1-a)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{26}$$

Suppose that the production function of good Y_d location j with the new technology PC is equal to

$$Y_d = \mathrm{PC}^{(1-\alpha)}[bH^{\rho} + (1-b)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{27}$$

where $a < b < 1$, which are personal computers. The firms need to decide the optimal amount of PC they want to pick. However, the decision of how much PC to choose increases with $\frac{H}{L}$. Before the availability of the PC technology, location j that had higher supply of skilled labor also had relatively low wages (because of a congestion effect on skills). Therefore, the return to skill increases most in locations that choose to adopt PC most intensively. However, the relationship between skill supply and return to skill is weakly decreasing. After the arrival of the PC technology, the relationship between the supply of skill and the return to skill is given by

$$\ln\frac{W_H}{W_L} = \begin{cases} \ln\left[\frac{aH^{\rho-1}}{(1-a)L^{\rho-1}}\right] & if\ \frac{H}{L} \leq \phi^L \\ \ln\left[\frac{a\phi^{L\,\rho-1}}{(1-a)}\right] = \ln\left[\frac{b\phi^{H\,\rho-1}}{(1-b)}\right] & if\ \phi^L < \frac{H}{L} < \phi^H \\ \ln\left[\frac{bH^{\rho-1}}{(1-b)L^{\rho-1}}\right] & if\ \frac{H}{L} \geq \phi^H \end{cases} \tag{28}$$

where ϕ^H and ϕ^L are the critical values of the skill ratio such that if a location is characterized by $\frac{H}{L} < \phi^L$, then it retains the old technology. If $\frac{H}{L} > \phi^H$, then the location switches to the new technology. Equation 2 shows that when a firm keeps the old technology, the relationship between the skill ratio and skill premium is negative, as if the firm had already switched to the new technology. However, when the firm is in transition between the old and new technologies, this relationship is equal to zero. This prediction of the model goes against fact 1 in figure 3. In fact, in figure 3, the relationship between the supply of skills and the skill premium becomes positive in the decade after 2000 and, overall, there is a positive trend. Therefore, a model with exogenous technological innovation seems better able to describe the data. It could also be the case that in order to obtain a positive relationship, I need a model that combines technological adoption and endogenous agglomeration forces.

D.3 Rewriting Y_T

In order to estimate the needed parameters, I compute the unobserved changes in cities' productivities, given the parameters of labor demand $\{\rho,\ \gamma^H,\ \gamma^L,\ \phi^H,\ \phi^L\}$ and the data $\{w_{Hjt},\ w_{Ljt},\ L_{jt},\ H_{jt},\ L_{djt},\ H_{djt}\}$. In order to make this transformation, I follow Diamond (2016) by taking the ratio of highly skilled wages to less skilled wages in location j:

$$\frac{w_{Hjt}}{w_{Ljt}} = \frac{\xi_{Hdjt} Y_{djt}^{1-\rho} H_{djt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H}}{\xi_{Ldjt} Y_{djt}^{1-\rho} L_{djt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (H_{jt} + L_{jt})^{\phi^L}} \implies$$

I use a change in the variable where defining highly skilled and less skilled productivities as

$$\xi_{Hdjt} = \theta(1 - \lambda_{djt})$$

$$\xi_{Ldjt} = \theta(\lambda_{djt})$$

This definition means that the skill premium can be written as:

$$\frac{w_{Hjt}}{w_{Ljt}} = \frac{\theta^{\frac{1}{\alpha}} (1 - \lambda_{djt}) Y_{djt}^{1-\rho} H_{jt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H}}{\theta^{\frac{1}{\alpha}} \lambda_{djt} Y_{djt}^{1-\rho} L_{djt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (H_{jt} + L_{jt})^{\phi^L}}$$

$$\frac{w_{Hjt}}{w_{Ljt}} = \frac{(H_{jt} + L_{jt})^{\phi^H - \phi^L}\ H_{jt}^{\gamma^H - \gamma^L} H_{djt}^{\rho-1} L_{jt}^{-\gamma^H + \gamma^L} (1 - \lambda_{djt})}{\lambda_{djt} L_{djt}^{\rho-1}}$$

$$w_{Hjt} L_{djt}^{\rho-1} \lambda_{djt} = (H_{jt} + L_{jt})^{\phi^H - \phi^L}\ H_{jt}^{\gamma^H - \gamma^L} H_{djt}^{\rho-1} L_{jt}^{-\gamma^H + \gamma^L} (1 - \lambda_{djt}) \implies$$

$$\implies \lambda_{djt}\left[w_{Hjt}L_{djt}^{\rho-1} + (H_{jt} + L_{jt})^{\phi^H-\phi^L} H_{jt}^{\gamma^H-\gamma^L} H_{djt}^{\rho-1} L_{jt}^{-\gamma^H+\gamma^L}\right] =$$
$$(H_{jt} + L_{jt})^{\phi^H-\phi^L} H_{jt}^{\gamma^H-\gamma^L} H_{djt}^{\rho-1} L_{jt}^{-\gamma^H+\gamma^L} w_{Ljt} \implies$$

$$\implies Y_{djt} = \left(\frac{(L_{jt}+H_{jt})^{\gamma pH} w_{Ljt} H^{\gamma^H} L^{-\gamma^L} H_{djt}^{\rho-1} L_{djt}^{\rho} + (L_{jt}+H_{jt})^{\gamma pH} w_{Hjt} H_{djt}^{\rho} H_{jt}^{\gamma^H} L_{djt}^{\rho-1} L_{jt}^{-\gamma^L}}{(L_{jt}+H_{jt})^{\gamma pH-\gamma pL} w_{Ljt} H_{djt}^{\rho-1} H_{jt}^{\gamma^H-\gamma^L} + L_{jt}^{\gamma^L-\gamma^H} + w_{Hjt} L_{djt}^{\rho-1} L_{jt}^{\gamma^L-\gamma^H}}\right)^{\frac{1}{\rho}}$$

This formulation of Y_{djt} is used in the estimation since it does not include the productivity terms S_H, S_L, ξ_H and ξ_L.

E Comparison with Spatial Equilibrium Models Without Technology Diffusion

One of the novelty of my model is the fact that, relative to the existing literature in spatial equilibrium, is embedded with convergence and divergence forces. The purpose of this section is to show how technological diffusion is key in introducing wage convergence into the model but also in matching the trends in the “Great Divergence” and the secular migration decline. Precisely, in this exercise, using the estimates generated in section 4.2.1, I simulate the model without the technological diffusion process. Table 14 shows how this version performs in terms of matching over time: (i) regional convergence pre-1980 and its decline, solely for highly skilled; (ii) the “Great Convergence” and the “Great Divergence” of skills and (iii) the decline in migration rates. Row (1) and (2) report the results on the overall wage convergence in the model and in the data. This model does not reproduce any convergence pre-1980 and the increase in divergence is very moderate. Overall, spatial equilibrium model have forces that help the sorting, therefore, increasing the divergence, but achieving convergence as we would see in the pre-1980 is quite hard. Therefore, in such model, we cannot explain the sharp changes observed in the data. Row (3) and (4) report the speed of the convergence rate for the highly skilled in the data and in the model, respectively. The model suggests that even before 1980, there would have bee divergence for highly skilled wages and this divergence would have increased over time. Similarly, row (5) and (6) suggest that data and model do not match either for the speed of regional convergence of the less skilled wages. Both groups look similar, despite the highly skilled wages show even more divergence coming from the larger agglomeration forces than the less skilled wages. Row (7) and (8) compare data and model prediction on the speed of convergence of the skill ratio, or the “Great Divergence”. The findings suggest that this version of the model cannot account for the “Great Convergence” pre-1980 and the decline of it is also very moderate as well. Finally, row (9)-(10) ((11)-(12)) compare the migration rates for highly (less) skilled. Interestingly, the findings suggest that migration across cities would have increased over time rather than decreased as in the data. This is because the model would produce extra sorting over time but there is no mechanism to slow down the incentive to move.

Overall, these results suggest that static spatial equilibrium models alone that do not have a growth component cannot match the sharp changes we have observed in the data in any of the salient moments.

Table 14: Model Fit with No Technological Diffusion

	1960-1980	**1970-1990**	**1980-2000**	**1990-2010**
β **Data**	-1.85	-1.50	-1.05	-0.11
β **Model**	0.04	0.05	0.06	0.06
β^H **Data**	-1.97	-1.35	-0.23	0.37
β^H **Model**	0.06	0.08	0.10	0.11
β^L **Data**	-1.78	-1.43	-1.96	-1.82
β^L **Model**	0.01	0.01	0.01	0.01
β^{HL} **Data**	-1.16	-0.14	-0.13	0.37
β^{HL} **Model**	-0.01	-0.02	-0.02	-0.03
% **Migr. Rate H Data**	4.44	3.62	3.33	2.17
% **Migr. Rate H Model**	23.51	33.44	37.16	45.35
% **Migr. Rate L Data**	2.23	2.05	2.26	1.19
% **Migr. Rate L Model**	22.62	30.41	32.49	36.30

F Online Data Appendix

In this subsection, I first describe in detail the data sets I use for the analysis. Second, I run several robustness checks for the decline in regional convergence.

F.0.1 Data Description

My two main data sets are the US Census data extracted from IPUMS. I use the 1% sample for 1940, 1% sample for 1950, metropolitan sample for 1970, 5% sample for 1980, 5% sample for 1990, and the 5% sample for 2000. Then, for 2010, I use information from the American Consumption Survey (ACS) extracted from IPUMS. I use information on wages, education, age, race, ethnicity, rents, birthplace, migration, population, industries, occupation, MSA, and state. All of this information is also available in the ACS data for 2010. I collect the same information from the CPS data set. The CPS is a monthly US household survey conducted jointly by the US Census Bureau and the Bureau of Labor Statistics. I use the observation for the month of March. The CPS data set is used mainly for the analysis of migration. My geographic unit of analysis is the MSA. An MSA is a "region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core." I also use two more data sets, one for the measure of Wharton land use regulation index (WLURI), aggregated by Saiz (2010) at the MSA level, and the other for the measure of RTI developed by Autor and Dorn (2013). The latter uses information on the task intensity of the occupation from the "O*NET" data set, which is available for download at http://online.onetcenter.org/.[30].

F.0.2 Robustness Checks

Before turning to the robustness tests, I provide one more time the specification for the β-convergence estimation that I use throughout the paper following the specification in 1. In most of the specifications, the observations are weighted by the initial size of the location j.

I run several robustness tests starting with the ones illustrated in figure 1 and in figure 2. I change the unit of analysis from cities to counties in figure 21. In figure 21, I plot the estimated convergence rates. In plot A, the estimate uses a 10-year rolling period, while plot B uses a 20-year rolling period. The convergence rate is negative and statistically significant until 1987 in plot A, while it is negative and statistically significant until 1997 in plot B. Both estimates show that the

[30]For a more detailed description of the RTI measure, please refer to Autor and Dorn (2013)

first period in which convergence ceased to be significant is 1978. This fact aligns with the findings of Higgins et al. (2006) who finds that there was convergence between 1970 and 1990. However, departing from this prior work, I conduct an analysis in which the period is extended and find that the convergence across counties follows the same patterns as the convergence across cities and states.

As a second robustness check, I show that the rate of convergence stops being significant and robust only if the initial year is after 1980. For this reason, I compute the rolling 20- and 30-year wage convergence as shown in figure 22 from 1940 onward. Then, I decompose it by skill group. Panels ((c))-((d)) and ((e))-((f)) of figure 22 show, respectively, results for the highly skilled and the less skilled groups. The rolling convergence rate β is negative and statistically different from zero until 1980, but then, it starts becoming positive but is still not significant. Finally, between 1990 and 2010, it becomes positive and statistically different from zero. But, when I decompose by skill groups, highly skilled workers show the same patterns as the aggregate convergence rate. Instead, the convergence rate for the less skilled group remains negative independently of the period. It actually becomes even stronger over time.

As a third robustness check, I reproduce figures 1 and 2 with compositionally adjusted wages. I control whether after using compositionally adjusted wages, the convergence rates change. As shown in figure 24, the convergence rates do not change substantially after adjusting for skill composition. Finally, another test is to see whether real wage convergence changes in the same way as nominal wage convergence. The caveat in looking at real wage convergence is that the data on local prices are very scarce, especially before 1980. For this reason, I use self-reported monthly rental prices as a proxy for local prices. As you can see in figure 23, real wage convergence decreases even more than nominal wage convergence after 1980. Specifically, decomposing by skill groups, the convergence rate is approximately zero in the less skilled group but becomes positive in the highly skilled group.

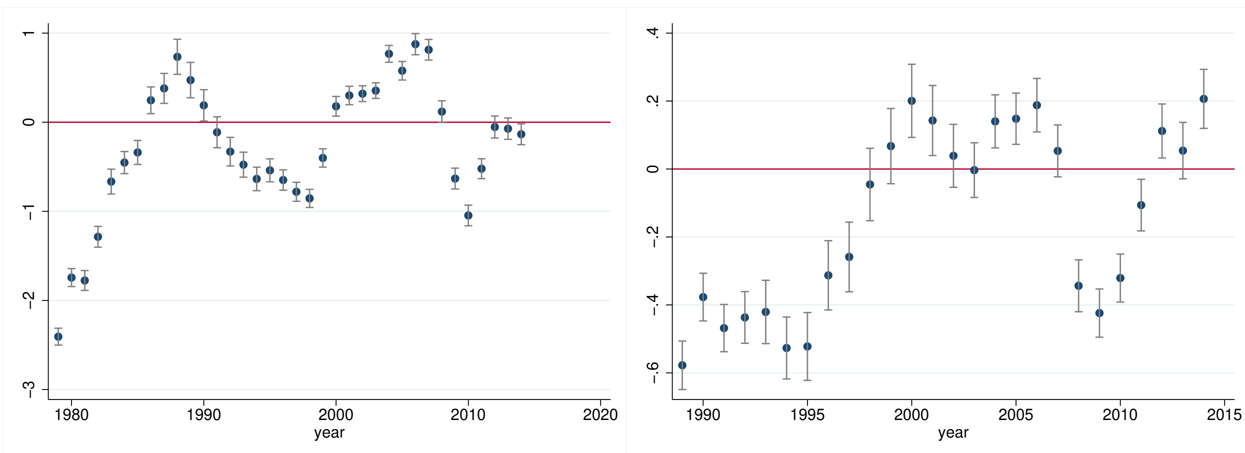
One reason why the convergence patterns might have changed could be because the definition of cities available between 1980 and 2010 is not perfectly identical to the one between 1940 and 1980. To make sure that it is not these different samples driving the slowdown in convergence, I estimate the unconditional cities' wage convergence between 1980 and 2010 by using the 127 cities available in 1940-1980. Table 16 shows the convergence rate after 1940 for the reduced sample. The results show that if I use only cities available before 1980, the convergence rate is even lower. Second, I look at the decline in wage convergence after adjusting for the skill-biased technical change shock. I run the following regression:

$$\Delta w_{jt} = \beta^o + \beta w_{jt-\tau} + \alpha^H \Delta S_{Hjt} + \alpha^L \Delta S_{Ljt} \tag{29}$$

where t is 2010 and tau is 30 years. After controlling for the technology shock, I get conditional convergence = -1.1% a year. This rate indicates that without taking into account the mechanisms of the model, SBTC affects the decline in wage convergence.

Figure 21: Convergence by county over time

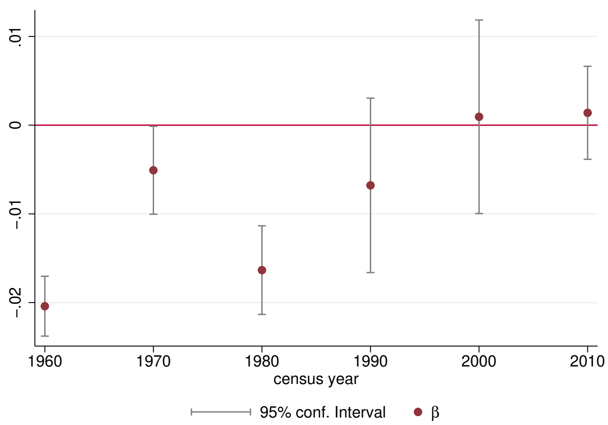
((a)) 10-year ((b)) 20-year



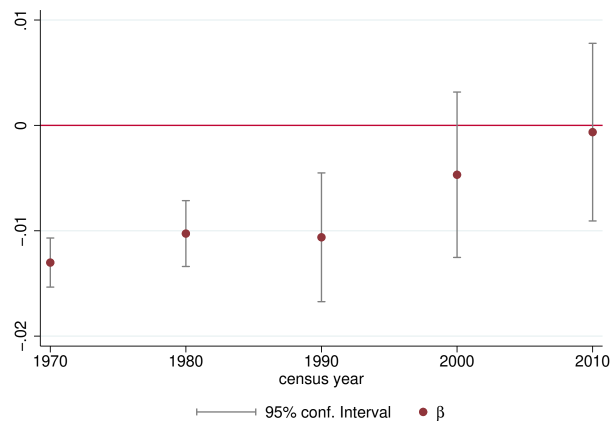
Note: Plot A shows the convergence rate at the county level for a 10-year rolling window that starts in 1969. Plot B shows the convergence rate at the county level for a 20-year rolling window that starts in 1969. Data for this analysis are from the Bureau of Economic Analysis Regional Economics Accounts. In each estimate the cities are weighted by their population. On the y-axis the coefficient is reported in percentage terms.

Figure 22: Convergence Rate Over Time - Overall and by Groups

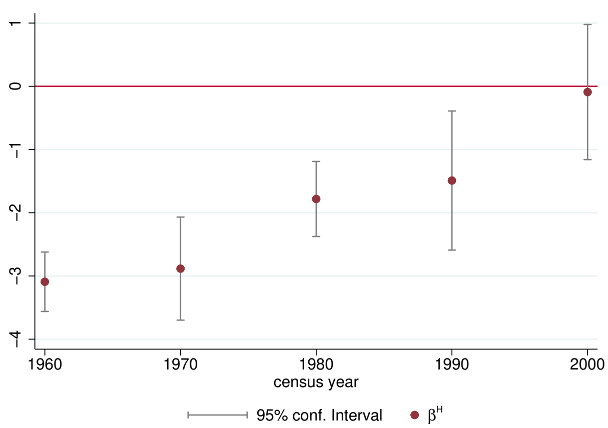
((a)) 20-year



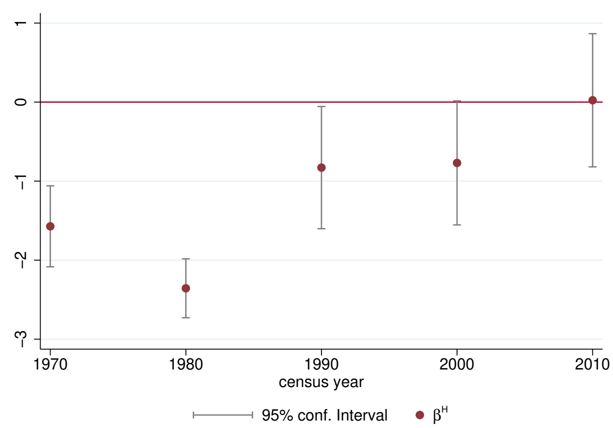
((b)) 30-year



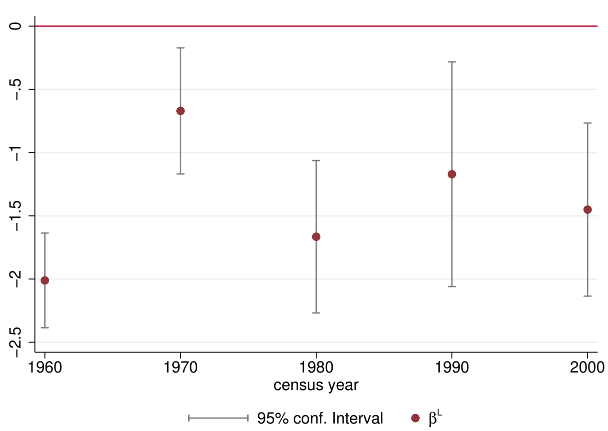
((c)) 20-year



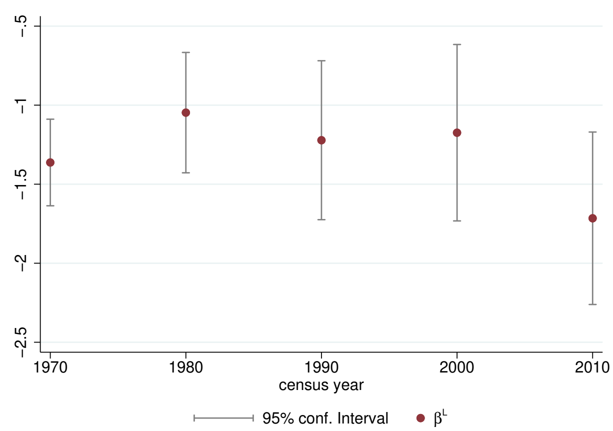
((d)) 30-year



((e)) 20-year



((f)) 30-year



Note: This figure shows the beta coefficient of the regression of the initial wage on the log wage changes using a 20-year and a 30-year rolling window. In each estimate the cities are weighted by their population. On the y-axis the coefficient is reported in percentage terms. Plots ((a)) and ((b)) are for the aggregate estimate of β, Plots ((c))-((d)) and ((e))-((f)) are, respectively, for β^H and β^L.

Table 15: Convergence Rates - Restricted Sample

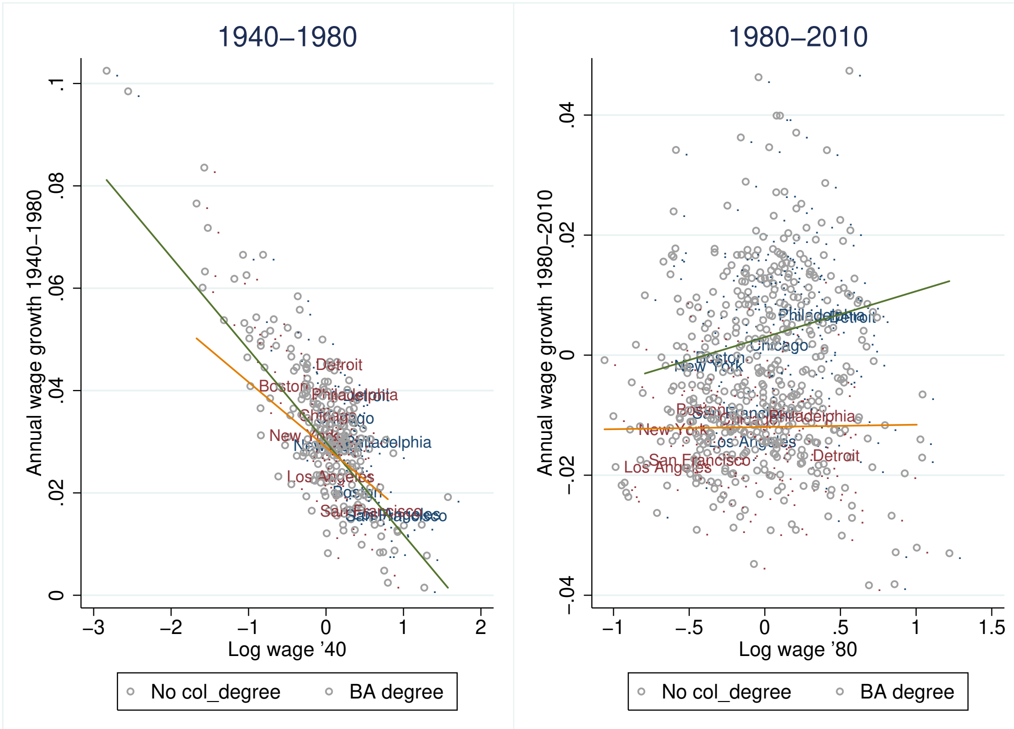
	(1) $\Delta^{1940-1980}$	(2) Δ_{-}^{80-08}
Log($wage^{1940}$)	-0.0109***	
	(-10.53)	
Log($wage^{1980}$)		-0.00116
		(-0.25)
Constant	-0.0217***	-0.0147***
	(-137.22)	(-24.45)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

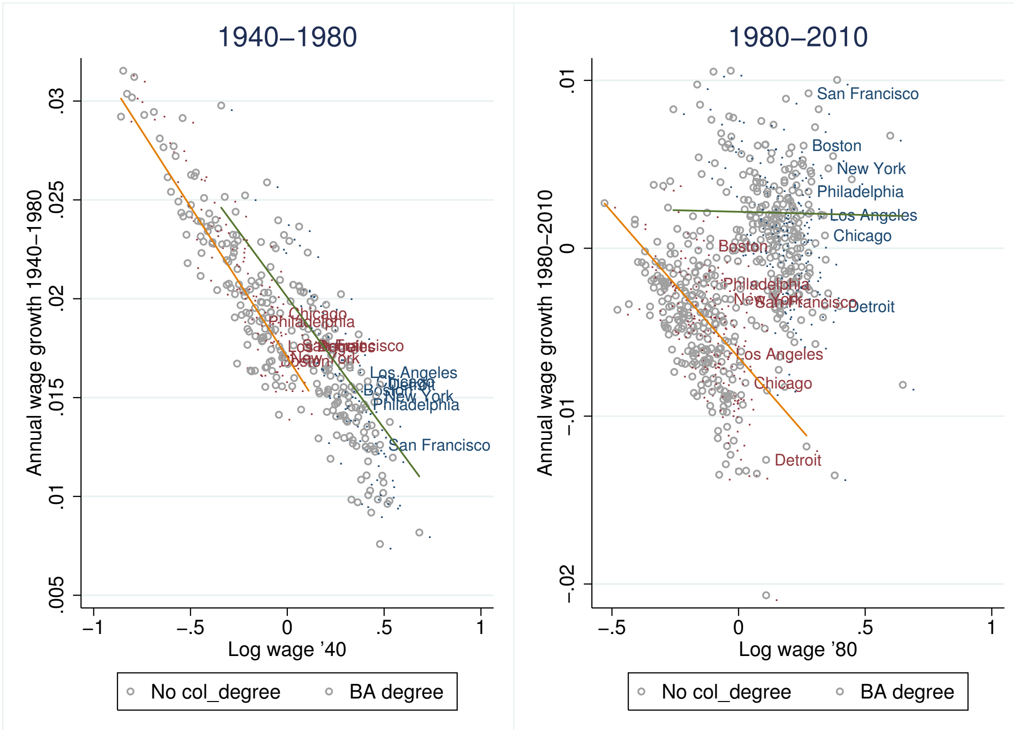
Note: I estimate the β convergence rate for the restricted sample with only 127 cities. In column (1), I estimate it for the 1940-1980 time period and in column (2) for the 1980-2010 time period.

Figure 23: Real Wage Convergence



Note: This figure shows two scatter plots of the log wages by MSA in the initial year against the annual average growth of the wages in the final year. The wages are divided by the rental prices in the MSA. The rental price is taken from the self-reported Census data. Specifically, on the left-hand side (right-hand side), I plot the demeaned log wages in 1940 (1980) by MSA against the annual average growth of wages between 1940 (1980) and 1980 (2010). The line in each graph represents a weighted regression line from the bi-variate regression.

Figure 24: Compositionally Adjusted Wage Convergence



Note: This figure shows two scatter plots of the log wages by MSA in the initial year against the annual average growth of wages in the final year. Wages are adjusted by individual characteristics, sex, race, age, marital status, before taking the MSA average. Specifically, on the left-hand side (right-hand side), I plot the demeaned log wages in 1940 (1980) by MSA against the annual average growth of the wages between 1940 (1980) and 1980 (2010). The line in each graph represents a weighted regression line from the bi-variate regression.

Table 16: Convergence Rates - Robustness

Panel A

	(1) 1940-1980	(2) 1980-2010	(3) 1940-1980	(4) 1980-2010
Log hourly wage, 1940	-0.0185***		-0.0189***	
	(-13.21)		(-12.99)	
Log hourly wage, 1980		0.00374		-0.00423*
		(0.96)		(-2.20)

Panel B

	(1) Δw^{40-80}_pw	(2) Δw_pw^{80-10}	(3) Δw^{40-80}	(4) Δw^{80-10}
Log($wage^{1940}$)	-0.0143***		-0.0164***	
	(-16.69)		(-26.63)	
Log($wage^{1980}$)		-0.00333		-0.0101***
		(-0.72)		(-3.76)

This table show the estimate of the β-convergence of the OLS. Columns (1) and (2) show the estimates, respectively, for 1940-1980 and 1980-2010 by using population-weighted observations. Columns (3) and (4) show the estimates, respectively, for 1940-1980 and 1980-2010 by using unweighted population observations. Panel A shows the estimates of the β-convergence for local wages adjusted by the rent in each MSA. Panel B shows the estimate of the β-convergence for compositionally adjusted wages.

Table 17: Convergence Rates by Skill- Robustness

	(1) No,'40-'80	(2) Yes,'40-'80	(3) No,'80-'10	(4) Yes,'80-'10
Panel A				
Log wage '40	-0.0127***	-0.0181***		
	(-7.01)	(-11.12)		
Log wage '80			0.000369	0.00764***
			(0.36)	(3.92)

	(1) No,'40-'80	(2) Yes,'40-'80	(3) No,'80-'10	(4) Yes,'80-'10
Panel B				
Log wage '40	-0.0203***	-0.0232***		
	(-13.82)	(-19.35)		
Log wage '80			-0.00425**	-0.00584*
			(-2.94)	(-2.36)

	(1) No,'40-'80	(2) Yes,'40-'80	(3) No,'80-'10	(4) Yes,'80-'10
Panel C				
Log wage '40	-0.0152***	-0.0133***		
	(-21.13)	(-11.78)		
Log wage '80			-0.0173***	-0.000381
			(-10.65)	(-0.19)

	(1) No,'40-'80	(2) Yes,'40-'80	(3) No,'80-'10	(4) Yes,'80-'10
Panel D				
Log wage '40	-0.0163***	-0.0202***		
	(-25.22)	(-19.86)		
Log wage '80			-0.0189***	-0.0104***
			(-11.96)	(-5.52)

Note: This table shows the estimate of the β-convergence of the OLS. Columns (1) and (2) show the estimates, respectively, for "No" college degree and for "Yes" college degree workers for the years 1940-1980. Columns (3) and (4) show the estimates, respectively, for "No" college degree and for "Yes" college degree workers for the years 1980-2010. Panel A has the estimates of the β-convergence by skill for local wages adjusted by the rent in each MSA. Panel B has the same estimates as in Panel A but the observations are not weighted by local population. Panel C has the estimate of the β-convergence for compositionally adjusted wages. Panel D has the same results but the observations are not weighted by MSA population.

F.1 More Empirical Evidence on the workers' skills, wages and migration premium

Fact: Migration Premium negatively correlated with wages of local pre-1980, positively correlation afterwards.

Migration Premium I define a new variable that I call the migration premium. In a nutshell, the migration premium is the difference between the wages of the migrants and the wages of the locals in a specific year and in a specific location. As above, I define migrants as all the workers who moved within the last year and locals the ones who did not. For the worker to be a migrant, he or she needs to have changed state in the last year. I compute the average of the compositionally adjusted wages for the workers who changed their state. Then, I compute the average of the compositionally adjusted wages for the workers that were already residing in that state before the previous year.

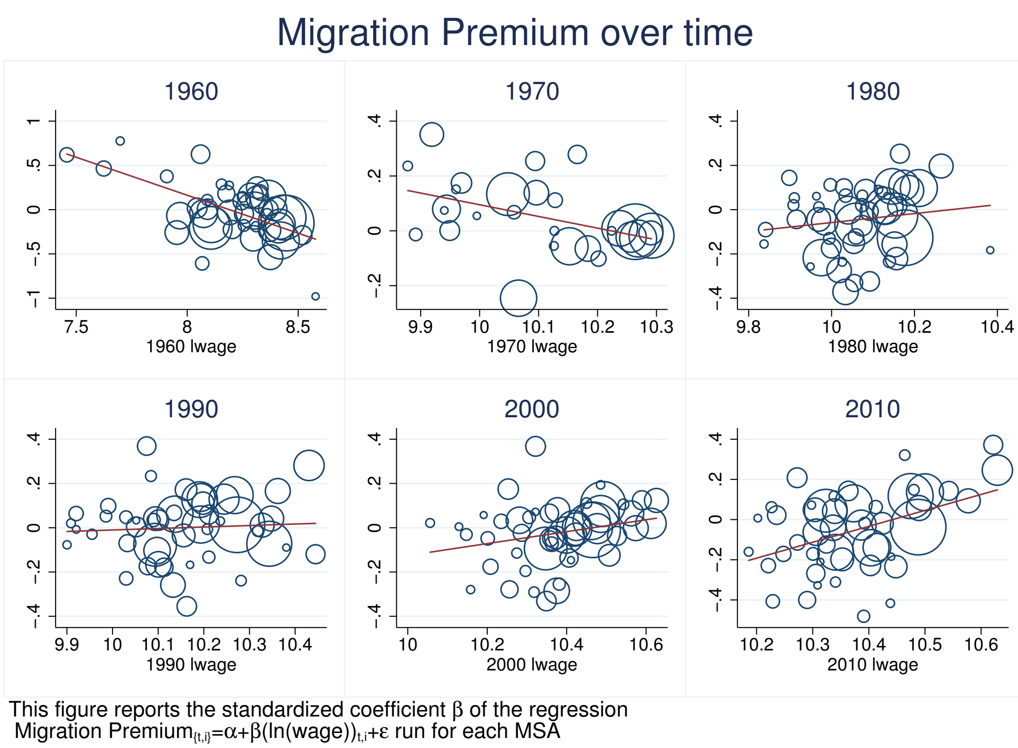
In figure 25, I look at the migration premium over time across states. For each of the years in the CPS sample, I run the following specification:

$$\ln\left(\frac{\hat{w}_{jt}^{migrant}}{\hat{w}_{jt}}\right) = \alpha_t + \beta_t \ln(\hat{w}_{jt}) + \epsilon_t$$

I run this specification for all the years of the sample in which the information on migration is available on CPS. Each regression is weighted by state population. Notice that the same results hold also for population.

In figure 25, the migration premium is defined as the difference between the wages of the migrants and the wages of the locals. The migration premium reported in figure 25 is adjusted for age, sex, race, nativity, and marital status. This figure shows that the migration premium is negatively correlated with the wage level of the state while the relationship becomes positive in 1980. I interpret this empirical finding as showing that the advantage of migrating until 1970 was higher in poorer statesmpirical contribution of this paper sh. While, later it became higher in the richer states.

Figure 25: Migration Premium by State over Time



Note: This figure shows a scatter plot of the log of the wages in the state in the first period t against the migration premium based on the measure of the difference between the wages of the migrants and the wages of the locals for the same year. The size of the underlying state is represented by the size of the circle in the figure. The line represents a weighted regression line from the bi-variate regression.