Skill-Biased Technical Change and Regional Convergence*

Elisa Giannone†

May 24, 2018

Abstract

Between 1940 and 1980, the wage gap between poorer US cities and richer ones shrank at an annual rate of roughly 1.4%. After 1980, however, there was no further regional convergence overall. In this paper, I quantify the contributions of skill-biased technical change and agglomeration to ending cross-city wage convergence within the US between 1980 and 2010. I develop and estimate a dynamic spatial equilibrium model that looks at the causes of regional convergence and divergence. The motivation behind my model choice is novel empirical regularities regarding the evolution of the skill premium and migration patterns over time and across space. The model successfully matches the quantitative features of the US regional wage convergence. Moreover, among others, it also reproduces "The Great Divergence" of skills across US cities, the secular decline in migration, and the increase in wage dispersion. Further, the counterfactual analysis indicates that the interaction between skill-biased technical change and agglomeration explains much of the change in cross-city wage differentials.

*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, Alejandro Justiniano, Munseob Lee, Kerstin Holzheu, Greg Kaplan, 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, and Maryland. 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.

†International Economics Section, Princeton University. Email at: giannone@princeton.edu

1 Introduction

Technological innovation was a major shock to the labor market toward the end of the twentieth century. It increased the wage differential between educated and less educated workers at the national level. At local level, for the majority of the twentieth century, poorer US cities had grown faster and were catching up with richer cities. However, in the 1980s, simultaneously with the technological innovation shock, cities with a lot of highly educated workers like Boston, San Francisco, and New York began to diverge. And this pattern was driven mostly by highly skilled workers. Although extensive literature has analyzed the impact of technological innovation at national level, curiously, little is known about its impact at regional level.

In this paper, I show that technological innovation contributed to the increase in the gap among US cities that had been closing. The key element in this phenomenon was the interaction between skill-biased technology and local agglomeration forces. Technological change gave an incentive to agglomerate where there was previously a higher concentration of skills. When highly skilled workers clustered together in these places, their wages rose even more. This made cities like Boston, San Francisco, and New York grow much faster than the others. Overall, these cities become amplifiers of the wage differences caused by skill-biased technological innovations.

Skill-biased technical change (SBTC) is the name assigned from the literature to the technological innovation that increased the relative demand for skills since 1980 (Katz and Murphy 1992). In turn, SBTC 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 in regional convergence and in the divergence of wages. In fact, at the regional level, between 1940 and 1980, wages in poorer US cities grew faster than wages in richer cities by 1.4% per year.[1] This wage convergence ended in 1980, and from 1980 to 2010 wages grew at similar rates in cities of different income levels. Figure 1 plots the annual average wage growth (demeaned) against its initial wage level in log (demeaned). The slope of the line, which estimates the β-convergence rate, is 0.014 between 1940 and 1980 but between 1980 and 2010, it goes to zero and is not statistically significant.[2]

The main contributions of this paper are threefold. First, I uncover novel empirical regularities regarding the evolution of regional convergence, skill premium and migration patterns over time and across space in the last 70 years. Second, I develop and estimate a dynamic spatial equilibrium model that analyzes the causes of regional wage convergence and divergence. I find that quantitatively, the model successfully matches the trends in wages among highly skilled (college educated) workers as well as the increase in spatial wage dispersion. Moreover, it also reproduces features of

[1] Notice that in the introduction, I use "cities" to refer to "Metropolitan Statistical Areas", which are my unit of geography. A definition is provided in section 2.

[2] 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.

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 size of each city's circle is proportionate to its initial population. The red line depicts a weighted least square bi-variate regression. Data come from the 2010 US Census and the 2010 American Community Survey.

the data on quantities such as "The Great Divergence" of skills in the last 70 years and the secular decline in migration. Third, I quantify the contribution of SBTC and agglomeration to ending cross-city wage convergence in the US in the last 30 years. This counterfactual analysis shows that the interaction of agglomeration and SBTC explains approximately 80% of the decline in regional convergence between 1980 and 2010.

The paper consists of three parts. In the first part, I identify a new set of facts about the evolution of wages and quantities by skill across cities during the last 70 years. First, I show that the end of wage convergence occurred only for highly skilled workers. Prior to 1980, the wage convergence rate for highly skilled workers and less skilled (non-college educated) workers was the same. Since 1980, the wages of less skilled workers have continued to converge at 1.4% annually, while the rate of wage convergence for highly skilled workers is 0%. Thus, any account of the end of convergence must distinguish between skill groups. Second, I show that in the last two decades, the relative price of skill has become positively correlated with relative quantities across cities. Specifically, after 1980, the correlation between the college ratio and the skill premium is positive across cities. Third, I show that since 1980, migration destinations of highly skilled workers have shifted towards already skill-abundant cities. These facts indicate that performance differences between highly skilled and less skilled workers played a crucial role in the cessation of regional wage convergence. Due to differences in their initial skill composition, some cities benefited more from SBTC. These facts are also consistent with the idea that demand forces become stronger than supply forces at the local level and then push the wages of highly skilled workers up more in cities where their concentration is higher.

Motivated by the new set of facts, in the second part of the paper, I develop a novel theoretical framework with skill-biased productivity shocks, local agglomeration spillover, and selective location decisions. The model nests two streams in the literature. First, in the spirit of Rosen (1979), Roback (1982), and Diamond (2016), I model local labor markets in which workers, highly skilled and less skilled, sample locations to live in that are heterogeneous along wages, rents, local amenities, and population in each sample period. Local markets are different in worker productivity depending on the agglomeration effect of population and skill concentration and in their exogenous productivity. Second, I follow the literature pioneered by Desmet and Rossi-Hansberg (2014) and Desmet et al. (2015) to introduce a dynamic component to a geography model. As I find, neither of these two models alone can account for regional trends in wages and in employment based on the agents' skills. These models falls short because the spatial equilibrium model lacks dynamism and the growth model lacks heterogeneity.

The model here allows trends in either direction. Convergence forces enter through a technology diffusion process as in Desmet and Rossi-Hansberg (2014). The key divergence forces are SBTC and agglomeration. To match the data and to account for other potential arguments, I also introduce other divergence forces such as costly migration and housing.[3]

I use a dynamic framework in general equilibrium with a final tradable sector, a set of intermediate inputs, and housing. Agglomeration affects all the intermediates equally. The productivity of the workers is skill-biased in two ways, an exogenous skill-biased force and an endogenous component that depends on the skill concentration of the city and its population. Households decide where to live and when to move, and they have a permanent loss in utility. They also decide how much to consume of the tradable and non-tradable goods.

The interaction of SBTC and agglomeration economies means that more educated locations have a larger skill premium. Highly and less skilled workers have some degree of complementarity, so, agglomeration effects raise the wages of all the workers. The differential increase in the wages of highly skilled workers makes the migration patterns for highly and less skilled workers diverge: highly skilled workers migrate to educated cities more than less skilled workers do. Migration has a twofold effect. First, if more workers migrate to a location, the marginal productivity of each

[3] Ganong and Shoag (2017) propose a mechanism based on housing regulations to explain the decline of wage convergence. Thus, I compare how much convergence there would be in the model if I did not account for housing.

decrease and hence, the returns decrease. Second, when more highly skilled workers move to a location, productivity goes up because of agglomeration effects, which raises the wages of all the workers especially those of the highly skilled workers.[4]

Therefore, technological innovation that is interacted with agglomeration forces counterbalances the convergence forces that were driving regional convergence before 1980. If technology were not skill-biased, the convergence forces would favor the poorer cities by pushing them toward the productivity frontier.

In the third part of the paper, I apply the model to the data. For the quantitative application, I build on Autor and Dorn (2013) to measure SBTC at the city level.[5] Autor and Dorn (2013) analyze the effect of computer innovation on the differences in the outputs of regional labor markets. Specifically, the arrival of computers mostly affects those occupations that are very routinized because machines can replace workers. Therefore, the effect of computers is heterogeneous across locations depending on how many highly skilled and less skilled workers hold very routinized occupations. Moreover, the industry's structure of routine intensive occupations can only predict the degree of specialization in the local labor market 10 years out. This limitation is the motivation for using a city's composition of routine intensive occupations and industries to capture the movements in the productivity of the workers and in the labor demand of the firms. The identifying assumption is that SBTC shocks, land unavailability, and the housing regulations are orthogonal to changes in local productivity. Following a similar identification strategy as in Diamond (2016), I structurally estimate the model using a GMM estimation procedure. I create moment conditions using equilibrium conditions from the model and local shocks that capture the skill-biased productivity component interacted with housing elasticities. I estimate the elasticity of population with respect to wages, which the prior research has not done. I find that an increase in a city's population raises the wages of both highly skilled and less skilled workers. I then supplement the estimates from the model with others borrowed from the literature to calibrate the model and solve it numerically.

The model has a good fit with the patterns of wage convergence in the last 70 years. In particular, it matches very closely the end of the wage convergence for the highly skilled workers after 1980 and the continuation of wage convergence 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 80% 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.

The model also matches other non-targeted moments closely such as the increase in wage dis-

[4] The definition of skill premium that I use is the difference between the wages of the highly skilled workers and those of less skilled workers.

[5] 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.

persion across cities between 1964 and 2009 in Hsieh and Moretti (2015). Moreover, through the lens of the model I show that the increase in wage dispersion across cities in the last 40 years happened solely among highly skilled workers. Hsieh and Moretti (2015) show that the increase in spatial wage dispersion is due to restrictive housing regulations in highly productive cities, like New York and San Francisco, in a context where they compare housing and amenities. I compare how much the increase in spatial wage dispersion is due to housing, migration costs, agglomeration, or SBTC. The results show that SBTC explains the biggest share of the increase in spatial wage dispersion. This result is novel and supplements the findings of Hsieh and Moretti (2015).

Besides the matching features of wages, the model matches the skill ratio. Notably, it shows that while there was convergence in the skill ratio before 1980, afterwards there was divergence, which Moretti (2012) calls "The Great Divergence."

My work contributes to the literature in several ways. The recent works that most closely relate to this paper study the increase in the spatial dispersion of US cities and states such as Hsieh et al. (2013), Hsieh and Moretti (2015), Ganong and Shoag (2017) and Baum-Snow et al. (Forthcoming). My paper complements these findings by highlighting the importance of skill-biased technology that is interacted with local agglomeration forces to explain the decline in regional convergence and other patterns.

My paper also speaks to the literature on the convergence and divergence in skill across cities as in Berry and Glaeser (2005), Moretti (2012), Eeckhout et al. (2014), Diamond (2016), Fajgelbaum and Gaubert (2018). I contribute to this literature by finding novel empirical evidence on skill sorting of migrants over the last 70 years. Moreover, my model fits the patterns in the data about the convergence of skill before 1980 and the divergence after 1980.

Additionally, this paper complements the literature on regional convergence across countries and states that is inspired by the seminal works of Baumol (1986), Barro and i Martin (1992), and Barro and Sala-i Martin (1997); and continued with Bernard and Jones (1996), Caselli and Coleman (2001), Comin and Ferrer (Forthcoming), and Gennaioli et al. (2014). My paper complements this literature in several dimensions. First, I provide a realistic model, rather than a North-South model. Second, I propose a framework that has both convergence and divergence forces built in that can match the data both on prices and quantities.

A growing literature, beginning with Katz and Murphy (1992) and continued by Krusell et al. (2000), has considered the impact of skill-biased technology on wages and inequality. Other works such as Berry and Glaeser (2005), Beaudry et al. (2010), and Autor and Dorn (2013) are more closely related to my paper because along with considering skill-biased technology as a national shock, they also focus on its local implications. In particular, Acemoglu and Autor (2011a) suggest using the difference between the definition of "tasks" and "skills" as a principal method to capture job polarization when discussing the skill premium.

This paper is also related to the agglomeration literature based on Krugman (1991). More recent papers in this literature are Greenstone et al. (2010), Behrens et al. (2014) and M. et al. (2015).

Duranton and Puga (2004) and Davis and Dingel (2014) propose, instead, a micro-foundation of agglomeration economies of skill. This paper complements the estimation and the long-run role of spatial agglomeration. In specific, it fits in the subset of this literature that makes use of spatial equilibrium models in general equilibrium to account for agglomeration forces (e.g., Redding and Rossi-Hansberg 2017, Desmet et al. 2015, Nagy 2016).

Methodologically, this paper relates to the literature that uses exogenous variation in local productivity to identify effects of labor demand as in Diamond (2016), Autor and Dorn (2013), Notowidigdo (2011), and Suárez Serrato and Zidar (2016). Overall, this paper complements all the papers above by: 1) the exploration of novel facts regarding the differential ending of convergence by skill groups, wage inequality at the regional level, and migration destinations over the last 70 years; 2) the incorporation and estimation of agglomeration effects of skill and size in a dynamic general equilibrium model and their quantification with heterogeneous skills; and 3) the introduction of a framework that can reproduce regional wage trends by skill group and could be used to explore other research questions.

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 calibrate the model that estimates the core parameters. In Section 5, I solve the model and conduct a counterfactual analysis. Section 6 explores other potential complementary channels. Section 7 concludes with a brief summary and future directions.

2 Data and Empirical Regularities

In this section, I briefly discuss the data and I report some facts for the last 70 years.

2.1 Data

Large sample sizes are important for a detailed geographical analysis of changes in the local labor markets. 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)).[6] 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.

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

2.2 The End of Wage Convergence for Highly skilled Workers after 1980

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.

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

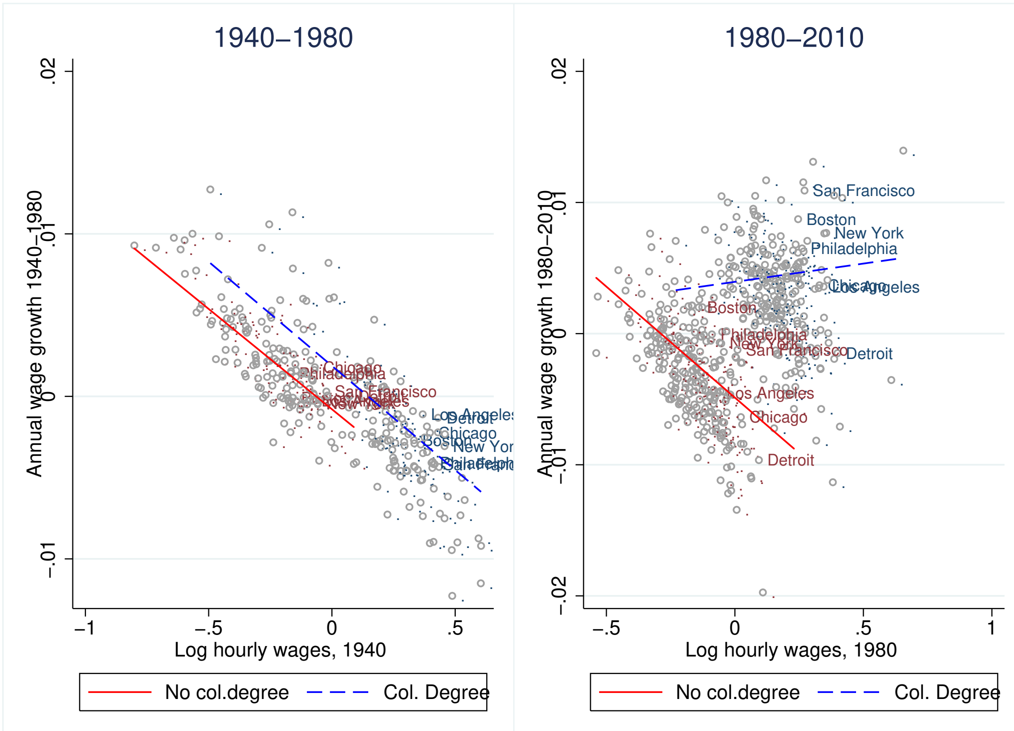
$$\frac{w_{kjt} - w_{kj\tau}}{(t-\tau)} = \alpha + \beta_k w_{kj\tau} + \epsilon \tag{1}$$

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 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 slope 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 the 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 the less skilled workers separately for 10- and 20-year windows. Second, I run the same regression as above for compositionally adjusted wages.[7]

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

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



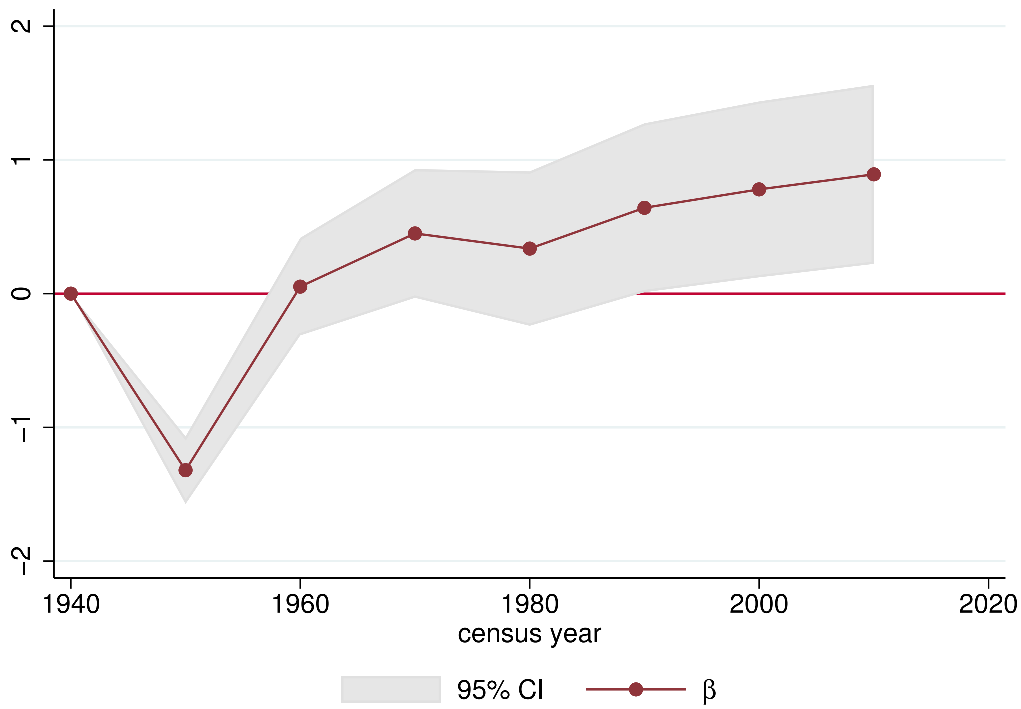
Note: This figure plots each MSA's annual average wage growth (demeaned) against its (demeaned) initial wage level by skill type (highly skilled and less skilled workers). The left depicts 1940-1980; the right depicts 1980-2008. Each MSA's circle size is proportionate to its initial population size by 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. The size of the underlying MSA is represented by the size of the circle in the figure.

The skill premium is lower in skill abundant places, and in recent years the skill premium is higher in skill abundant places as shown 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.

2.3 Migration

In this subsection, I use the CPS data. This data set has better migration information than the IPUMS extracted Census data. The information is reported more frequently than the IPUMS extracted Census data. Moreover, the information is more detailed and shows whether the workers migrate across MSAs, not only states. as the IPUMS extracted Census data do. As a robustness check, I run the same exercise with the migration measure computed as the difference between the individuals born in a state minus the individuals currently living there. The results in qualitative terms stay the same.

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).

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

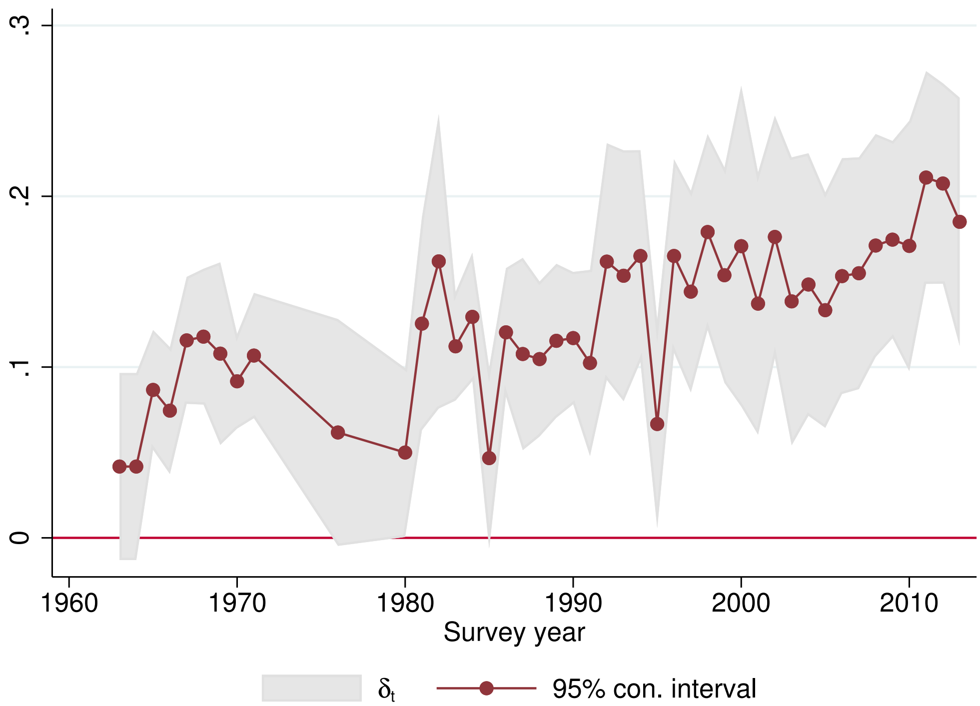
The dependent variable in this equation is whether worker i in MSA j at time t is a migrant or not. The variable equals one if the worker is a migrant. On the right-hand side, there is an indicator variable H_{ijt} that equals one if the worker is highly skilled and zero otherwise. The second variable 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.[8]

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 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.

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

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.

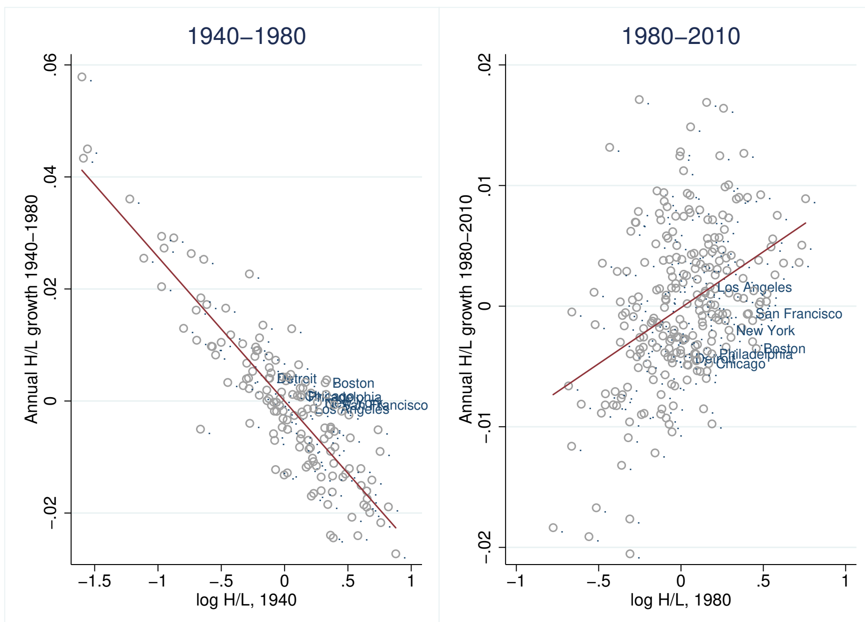
2.4 Skill Ratio: The Convergence and "The Great Divergence" after 1980

What happened to the distribution of highly and less skilled workers over time across space? Moretti (2004), Berry and Glaeser (2005), Diamond (2016), and Moretti (2012) 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 1970? Was the skill distribution converging across MSAs when wages were converging? To answer this question, I look at the convergence rates of the skill ratio over the last 70 years, not just between 1970 and 2000 as in the literature. 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 16, 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. However, as suggested in the literature, 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. In particular, a 1% increase in the college ratio in 1950 decreases the change in the college ratio between 1980 and 1980 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 1980-2008. Each MSA's circle size is proportionate to its initial population size. The red line depicts a weighted least square bi-variate regression. The size of the underlying MSA is represented by the size of the circle in the figure. The line in each graph represents a weighted regression line from the bi-variate regression.

3 A dynamic spatial equilibrium model with heterogeneous skills

The empirical analysis indicates that the skill composition might explain the patterns observed in the skill premium and cross-MSA migration. It also emphasizes the way in which the role of the skill composition changes over time. These observations also show that agglomeration and SBTC effects might be relevant in explaining why spatial wage convergence decreases. But how can these effects be disentangled? How important is each of these mechanisms? To answer these questions, I build a dynamic model of cities based on the current spatial equilibrium literature, which was pioneered by Rosen (1979) and Roback (1982), and is nested with the literature on economic growth. Specifically, this framework uses several features from Diamond (2016) and Desmet et al. (2015). It departs from Diamond (2016) in several aspects. First, I add a dynamic component resembling Desmet et al. (2015). Second, I add a set of intermediate sectors. Third, I emphasize the agglomeration effects on the production side of the economy by disentangling population and skill elasticities. I allow the prices of the rents to change across space for the composite good. This flexibility means I can obtain a series of rents and local prices after 1940.[9] Fourth, I shift the focus away from the micro aspects of location choice, such as distance from the native state and local preferences for amenities. I also do not estimate the model for multiple groups.[10] At the same time, this model borrows its dynamic structure from Desmet et al. (2015). However, it departs from the latter by adding heterogeneous agents, intermediate goods, and a housing sector.

There are two types k of households, highly skilled H and less skilled L. In each time period t, where $t \in T$ they decide how much to consume and which location j to pick for living. There is a set of J 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 he or she lives for which he or she is compensated with a wage. 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 production is a reduced form of the prices.

3.1 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 $\bar{O}$. Agents also derive utility from

[9] This work contributes to the spatial economics literature that currently lacks a measure of local prices back in time.

[10] While Diamond (2016) focuses on welfare and on the heterogeneity of workers, this paper asks a long-run macro aggregate question.

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

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

where ζ is a taste shock distributed according to a Gumbell (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. (2015) 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.[11] 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

[11] Caliendo et al. 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

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

From the last line of equation 3.1, 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. Thus,

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

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

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

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

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

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

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

$$
H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))} \tag{5}
$$

$$
L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{6}
$$

where

$$\delta_{kjt} = \theta \ln(W_{kjt} - R_{jt}\bar{O}) + (1-\theta)[\ln((1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + A_{kjt} + \gamma^p \ln\left(H_{jt}/L_{jt}\right) \tag{7}$$

3.2 Technology

In the next subsection, I describe the production technology of the final tradable sector, T; the non-tradable intermediates; and the housing sector, O. The final good is produced using all the intermediates jointly in a Cobb-Douglas 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 nummeraire.

3.3 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. In particular,

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

3.3.1 Intermediates Sector

The production function in equation 8 is a CES that uses two types of labor H_{djt} and L_{djt} as imperfect substitute inputs.[12]

$$Y_{djt} = [\eta_{Ldjt} L_{djt}^{\rho} + \eta_{Hdjt} H_{djt}^{\rho}]^{\frac{1}{\rho}}, \qquad \forall\, j = \{1, ..., N\} \tag{8}$$

η_{Hdjt} and η_{Ldjt} denote the productivity of H and L, respectively, in sector d at location j for time t. Productivity is divided into an exogenous and an endogenous component.[13]

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

[12] 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.

[13] 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 B.3.

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.[14] 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.[15] It follows that the expressions for η_{Hdjt} and η_{Ldjt} are:

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

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

where S_{kt} is the exogenous skill-biased technology component for $k \in \{H, L\}$.[16] The exogenous productivity component is ξ_{kdjt}. ξ_{kdjt} at time 0 is given and then evolves according to:

$$\xi_{kdjt} = \xi_{kdjt-1}^{\gamma_2} \left[\int_s \omega(j,s) \xi_{kdst-1} ds \right]^{1-\gamma_2} \tag{9}$$

where $\omega(j,s)$ is a symmetric measure of distance between location j and location s and $\gamma_2 \in [0,1]$.[17] If $\gamma_2 < 1$, then the productivity in location j is dependent on the productivity of the other 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

[14] 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.

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

[16] 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.

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

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 model becomes a repeated static model, which facilitates numerical solution of the model.

Since the labor markets are perfectly competitive, the wage in each location is equal to the marginal product of labor as shown in equations 10 and 11, which derive the first-order condition of the firms.

$$W_{Hjt} = p_{djt}\eta_{Hdjt}[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}H^{\rho-1}_{djt} \tag{10}$$

$$W_{Ljt} = p_{djt}\eta_{Ldjt}[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}L^{\rho-1}_{djt} \tag{11}$$

3.3.2 Housing Market

The supply of housing is a convex function of its price. The higher the price of housing, the higher the supply.[18]

$$O_{jt} = R^{\mu}_{jt} \tag{12}$$

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

$$H_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{13}$$

3.4 Equilibrium

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

Definition The equilibrium consists of a set of allocations $\{\{L_{djt},\, H_{djt}\}^{D}_{d=1}\}^{J}_{j=1}$ and a set of prices $\{\{P_{djt}\}^{D}_{d=1},\, R_{jt}\}^{J}_{j=1}$, wages $\{W_{Hjt}, W_{Ljt}\}^{J}_{j=1}$, such that given $\{\{\xi_{Ldj0},\, \xi_{Hdj0}\}^{D}_{d=1}\}^{J}_{j=1}$, $\{\{A_{Ljt},\, A_{Hjt}\}^{T}_{t=1}\}^{J}_{j=1}$, 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 5, 6 and 7;

2. Firms maximize profits such that equations 10, 11 hold;

[18] To create 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.

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

4. Labor markets clear such that 5 and 6 hold;

5. Housing markets clear such that the demand of equation 13 is equal to the supply in equation 12
$$R_{jt}^{\mu} = H_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Ljt}}{R_{jt}}\right]$$

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 9.

3.5 Discussion

Introducing this persistent productivity formulation with spatial diffusion helps to generate convergence directly in the model, as in Barro and Sala-i Martin (1997), Caselli and Coleman (2001), and Desmet et al. (2015). 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) construct 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 might be related to a declining cost of human capital or to physical capital mobility, as in the literature.

The upside of the model used here is that it extends the existing spatial equilibrium models by adding a dynamic component. This addition allows me to view income convergence through the lens of technological diffusion rather than TFP residuals, as would be the case in a static model. Workers draw idiosyncratic shocks every period that incentivizes them to switch cities. In a model with no agglomeration forces, a positive number of workers would find it optimal to switch cities, and the model would have positive flows of workers across cities, even in the absence of productivity shocks. This feature of the model accords well with the data in which the flows of workers are always positive. Net flows increase as a fraction of workers move to the relatively more productive sector that then decreases the difference in utility values across cities. In the next period, when taste shocks are drawn again, there is still positive net reallocation, but the net reallocation declines as the difference in values across cities declines. This process continues until the new desired allocation is achieved and flows return to zero. However, while preference shocks act as a dispersion force as just described, if agglomeration economies are stronger, there could be multiple equilibria. Therefore, other congestion forces serve to match the data.

This model features labor as the only production input. There is no 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.

3.6 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. (2015) 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. (2015) 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.

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} an S_{kt} for $k \in \{H, L\}$. 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 $\{\theta, \gamma^p, \gamma^L, \gamma^H, \rho, \phi^H, \phi^L, \lambda^H, \lambda^L\}$ 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. In particular, 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 other parameters $\{m_{2H}, m_{2L}, \mu, \bar{O}, \gamma_2, \alpha, \mu_d, \forall d\}$ with data from the literature.

4.1 Estimation of the Model

To estimate the model, I extend it to allow S_{kt} for $k \in \{H, L\}$ to be location j specific that becomes S_{kjt} for $k \in \{H, L\}$. This extension allows for extra variation generated by the city to estimate the parameters. From the wage equations of the tradable sector, it follows that the exogenous change in productivity is divided into two main components. On the one hand, there are skill-biased productivity shocks ΔS_{Hjt} and ΔS_{Ljt} that act as divergence forces while, on the other hand, the other skill-neutral component ξ_{Ht} and ξ_{Lt} pushes poorer cities to reach the productivity frontier more quickly. An important assumption is the linearity between the technology component ΔS_{Hjt} and ΔS_{Ljt} and the exogenous productivity ξ_{Hjt} and ξ_{Ljt}.

4.1.1 Skill-Biased Productivity Shock

Autor and Dorn (2013) rank commuting zones by the intensity of the routine occupations.[19] 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 skilled and the 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 14 and 15.

$$\Delta S_{Ljt} = \sum_{\upsilon=1}^{\Upsilon} \left(\frac{L_{j\upsilon t}}{L_{jt}} - \frac{L_{j\upsilon t-10}}{L_{jt-10}} \right) 1\left(RTI_{\upsilon} > RTI_{P66}\right) \tag{14}$$

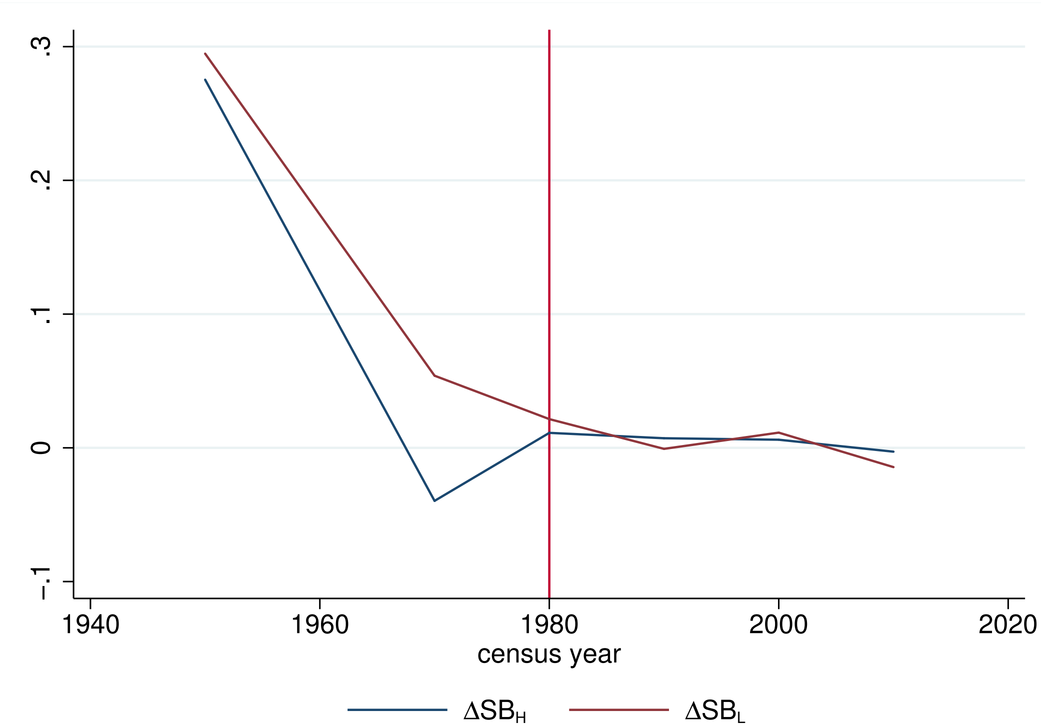
$$\Delta S_{Hjt} = \sum_{\upsilon=1}^{\Upsilon} \left(\frac{H_{j\upsilon t}}{H_{jt}} - \frac{H_{j\upsilon t-10}}{H_{jt-10}} \right) 1\left(RTI_{\upsilon} > RTI_{P66}\right) \tag{15}$$

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 the less skilled workers as in equations 14 and 15. ΔS_{Hjt} and ΔS_{Ljt} are two good proxies for how SBTC affects cities

[19] 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/

in different ways depending on their composition.[20] Figure 6 shows the evolution over time of the aggregate measures. However, this is not a good measure of a productivity shock because it

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



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

$$\Delta \tilde{S}_{Hjt-10} = \sum_d \left(H_{d-jt} - H_{d-jt-10}\right)\left(R_{djt-10}\right)$$

$$\Delta \tilde{S}_{Ljt-10} = \sum_d \left(L_{d-jt} - L_{d-jt-10}\right)\left(R_{djt-10}\right)$$

[20]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.

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 **??** 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. In particular, in the 1970s, the F-statistic is less than 7.[21]

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 a moment in order to estimate the set of parameters: $\{\gamma_H, \gamma_L, \phi^H, \phi^L, \rho, \lambda^H, \lambda^L\}$.

For this purpose, I start by taking the logs and the first differences of equations 11, 10, which, respectively, give:

$$\Delta w_{Ljt} = (1-\rho)\Delta \ln Y_{djt}(\rho, \gamma_{H,} \gamma_L, \phi_{H,} \phi_L) + (\rho - 1)\Delta \ln L_{djt} + \gamma_L \Delta \ln \frac{H_{jt}}{L_{jt}} + \\ + \phi^L \Delta \ln (H_{jt} + L_{jt}) + \lambda^L \Delta S_{Ljt} + \Delta\xi_{Ldjt}$$

$$\Delta w_{Hjt} = (1-\rho)\Delta \ln Y_{djt}(\rho, \gamma_{H,} \gamma_L, \phi_{H,} \phi_L) + (\rho - 1)\Delta \ln H_{djt} + \gamma_H \Delta \ln \frac{H_{jt}}{L_{jt}} + \\ + \phi^H \Delta \ln (H_{jt} + L_{jt}) + \lambda^H \Delta S_{Hjt} + \Delta\xi_{Hdjt}$$

As in 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. Also, the interaction of SBTC shocks with the cities' housing supply elasticities leads to variation in the

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

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 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:[22]

$$\mathrm{E}(\Delta\xi_{Hdjt}\Delta Z_{jt}) = 0$$

$$\mathrm{E}(\Delta\xi_{Ldjt}\Delta Z_{jt}) = 0$$

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

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

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

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 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. In particular, 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. In particular, the moment restrictions are:

$$\mathrm{E}(\Delta A_{Hjt}\Delta Z_{jt}) = 0$$

$$\mathrm{E}(\Delta A_{Ljt}\Delta Z_{jt}) = 0$$

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

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.

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. Following the literature, I can use the estimate from Notowidigdo (2011) that provides separate migration costs for highly and less skilled workers. Notowidigdo (2011) uses an exponential function to estimate migration costs. The functional form he estimates is as follows:

$$m_{k2j} = \frac{\sigma^k exp(\beta^k x_j) - 1}{\beta^k}$$

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.

4.2.1 Estimation Results

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 several specifications to test the sensitivity of the model. The results of the estimates, overall, are thus in accordance with the literature. I estimate an elasticity of substitution between highly and less skilled labor of 2.13 in column (1). In the other columns, the estimates vary from 1.82 to 3.5, which are consistent with the literature. The results related to endogenous spillover of skills, γ^H and γ^L, are similar to Diamond (2016)who uses similar functional forms. The estimates show that the returns to education are strong for highly skilled workers. In particular, I find that a 1% increase in the share of highly skilled workers raises their wages by 0.616% but increases the return for less skilled workers. Two novel parameters are ϕ^H and ϕ^L, the endogenous effect of population productivity. An increase in a MSA's population, on the other hand, does not affect the wages of highly skilled workers in column (1). But, columns (2)-(6) indicate that a 1% increase in a city's population decreases the highly skilled workers' wages between 0.262% and 0.646%. These estimates are similar to the one in Diamond (2016) who finds a value of 0.647%-0.697%. In column (1), the effect on the wages of less skilled workers is estimated to be negative such that a 1% increase in the city's population decreases their wages by 0.111%. In columns (2)-(6), the effect is reversed and a 1% increase in the city's population decreases their wages by 0.142%-0.187%. These estimates do differ slightly from Diamond (2016) who finds that a larger number for the wages of less skilled workers in absolute value. Similar estimates are also conducted by Moretti (2004) who finds that a 1% increase in MSA's college employment ratio leads to a 0.10% increase in the wages of highly skilled workers and a 0.16% increase in the wages of less skilled workers. 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. My results are well in accordance with their findings. I also report estimates of the labor supply. The share of tradable T versus the share of housing in the utility function of the workers, θ, varies from 0.468% and 0.577%. This range indicates that about half of the expenditures are on housing. Overall, these estimates show that workers, in general, prefer cities with higher wages, lower rents, higher college share. A 1% increase in the skill ratio increases the local highly and less skilled working population by 0.679%. This value is lower than that in Diamond (2016) and Albouy (2012). However, the periods and identification strategies used here are not the same as those used in these prior papers, which might be the reason why the value on housing is lower. In columns (3)-(6), I split the spillover effect of the college ratio on utility by skill group. I find that by splitting them, the effect is positive and statistically significant for highly skilled workers and ranges from 1.884% to 1.938% and is negative for low skill-workers going from -0.420% to -0.351%. I also report the estimate for the coefficient for the SBTC measure. This estimate serves as a scale of the effect. I report it separately for the case in which I estimate it as being the same for highly and less skilled workers. The estimates range between -0.125 to -0.014. However, when I separate the estimate for high and low, I find that the estimates tend to be larger for the high than for the low as reported in column (3). I run a test of the over-identifying restrictions to assess whether my instruments are jointly uncorrelated with unobserved local amenity changes and unobserved local productivity changes. I cannot reject the hypothesis that my instruments are jointly uncorrelated with unobserved local amenity changes and unobserved local productivity changes with p-values larger than 0.05 in columns (1) and (2).

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.

I borrow the parameter for the technology evolution process γ_2, which relates to the relationship between growth and population distribution, from Desmet et al. (2015). The parameters of the migration cost function, which is exponential, are different for highly skilled and the 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 Fast Marching Algorithm from Desmet et al. (2015). While Desmet et al. (2015) uses this procedure for one type of workers, the analysis could be extended to two types of workers.[23]

5 Model Simulation and Counterfactuals

In this section, I first describe how I do the estimation. Second, I provide more details as to how I achieve the numerical computation of the equilibrium. Third, 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. Fourth, I conduct a quantitative decomposition of each mechanism's effect on the decline in convergence. Fifth, 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.

The estimation procedure obtains the values for all 10 model parameters, the initial productivity terms, and 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 10, 11, 5, and 6. 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 found in appendix's section B.1.

5.1 Model vs. Data

Overall, the model provides a good fit to the patterns found in the previous subsections. Specifically, with the wages for highly and less skilled workers that the model produces, I run the same regression within my model as I did with the data in section 2. In fact, I also estimate in the model the β_k-convergence using the specification 1 proposed in section 2, which follows Baumol (1986). Then, I show that the model provides a good fit to the convergence patterns found earlier. I construct the evolution of the β-convergence for average wages and for wages of highly- and less skilled workers, which follows equation 1.

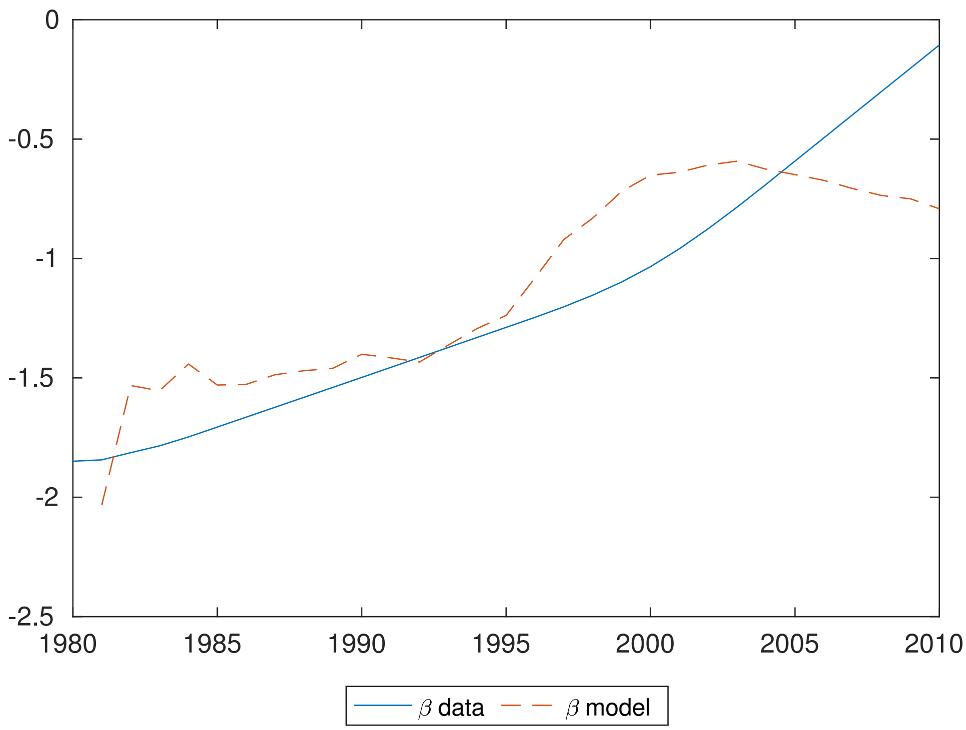
In Figure 7, I plot the estimated β-convergence from the model and from the data to compare them. I average out the estimates of β^H- and β^L-convergence weighted by the population shares. The estimates from the data look like a step function since the data are available only every 10

[23] 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.

years. Instead, the estimates from the model can be computed every year. Overall, the match is good. 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 well in fitting the wage convergence patterns in the non-targeted moment and the decline of convergence for the high(low)-skill group.

Figure 7: Model Matching the Data on Wage Convergence



Note: This figure shows a rolling estimate of the β-convergence over 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.

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 stepwise 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. Plot (a) of figure 9 shows the 10-year window β convergence starting with 1979 as first year, When

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 over 30 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.

I set γ^H, γ^L, ϕ^H, ϕ^L, γ^p to 0 after 1980, the change is quite dramatic suggesting that β would have increased substantially initially but it would have set around -1.3% in 2010. When I turn off also SBTC, as in plot (b), by 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. In plot (c) and (d), respectively, I 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 few 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. Plot (a) shows a similar pattern to plot (a) of figure 9 with an overall decrease in convergence ending at -.5% in 2010. In plot (b), when turning off also SBTC foirces, the overall convergence rate would have declined to -1.75%. In plot (c) and (d), respectively, I find that they move the convergence rate by only few percentage points as for β. In figure 11, I run the same exercise for β^L. In plot (a), I find that if agglomeration forces had been set to 0, then, overall, β^L would have decreased to almost -4% and finalized to -2.5% in 2010. However, in plot (b), when I also shut down SBTC forces, the exercise shows that convergence in the last 10-year window would have decreased more than in the baseline, approximately to -.5%. In plot (c) and (d), 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 demonstrate that the convergence rate with endogenous productivity channels and SBTC are about 1.2% a year. But, if I shut down the productivity channel, nominal wage convergence is about 1.1% a year.

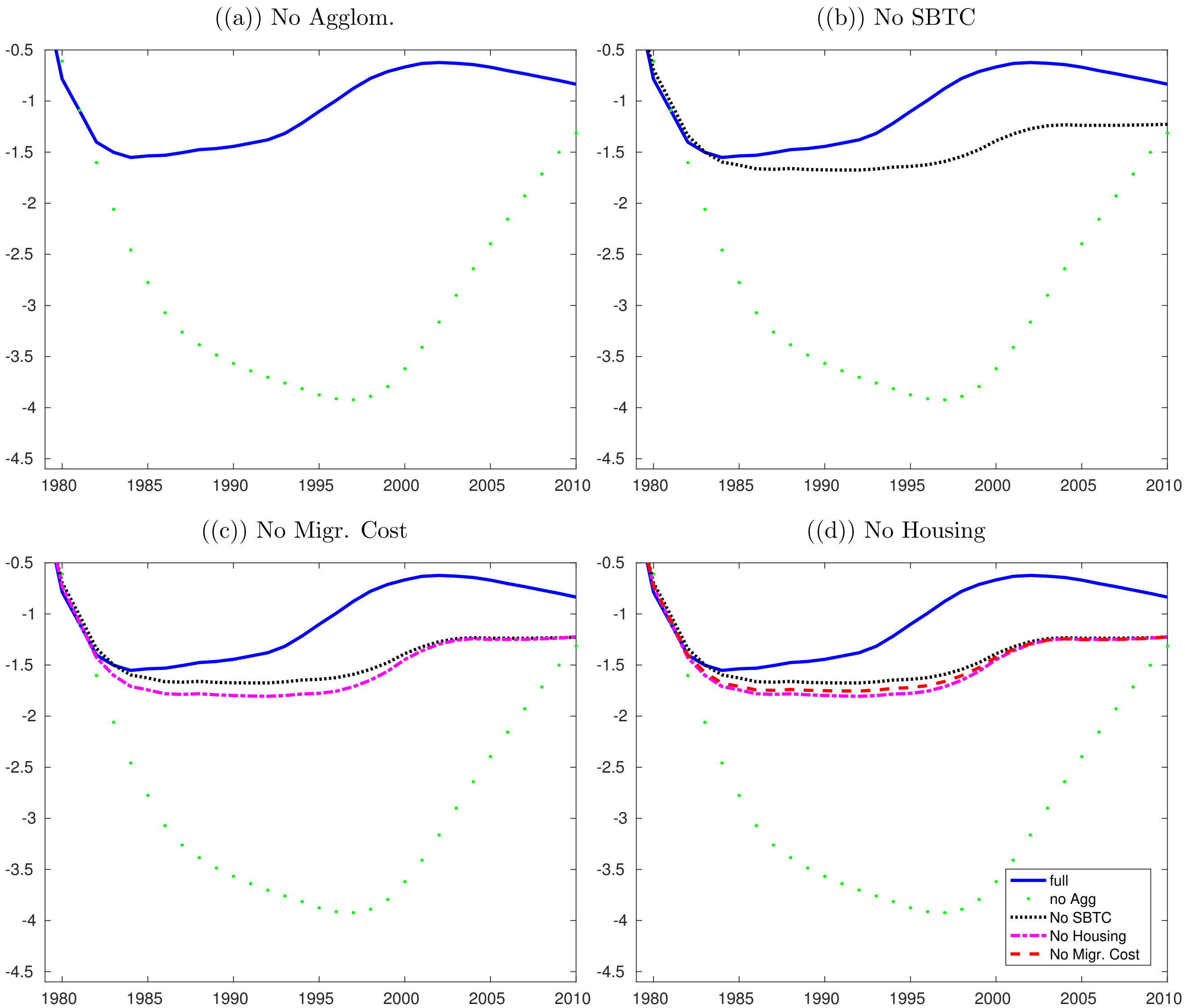
5.3 Wage Dispersion Increase Over Time

Hsieh and Moretti (2015) show that wage dispersion across US cities increased substantially between 1964 and 2009. As in table 11, 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 increased only in the highly skilled group but it decreased in the less skilled group. As shown in table 11, the variance between 1964 and 2009 increased by 519% in the highly skilled group. Instead, it decreased by 82% for the less skilled group. I run some counterfactual analysis by shutting down agglomeration forces as in section 5.2 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 19% for the highly skilled as shown in column (2). In column (3), I also shut down the SBTC finding that variance would have decreased both for the highly and for the 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.

5.4 The Convergence and Divergence of Skill ratio over time

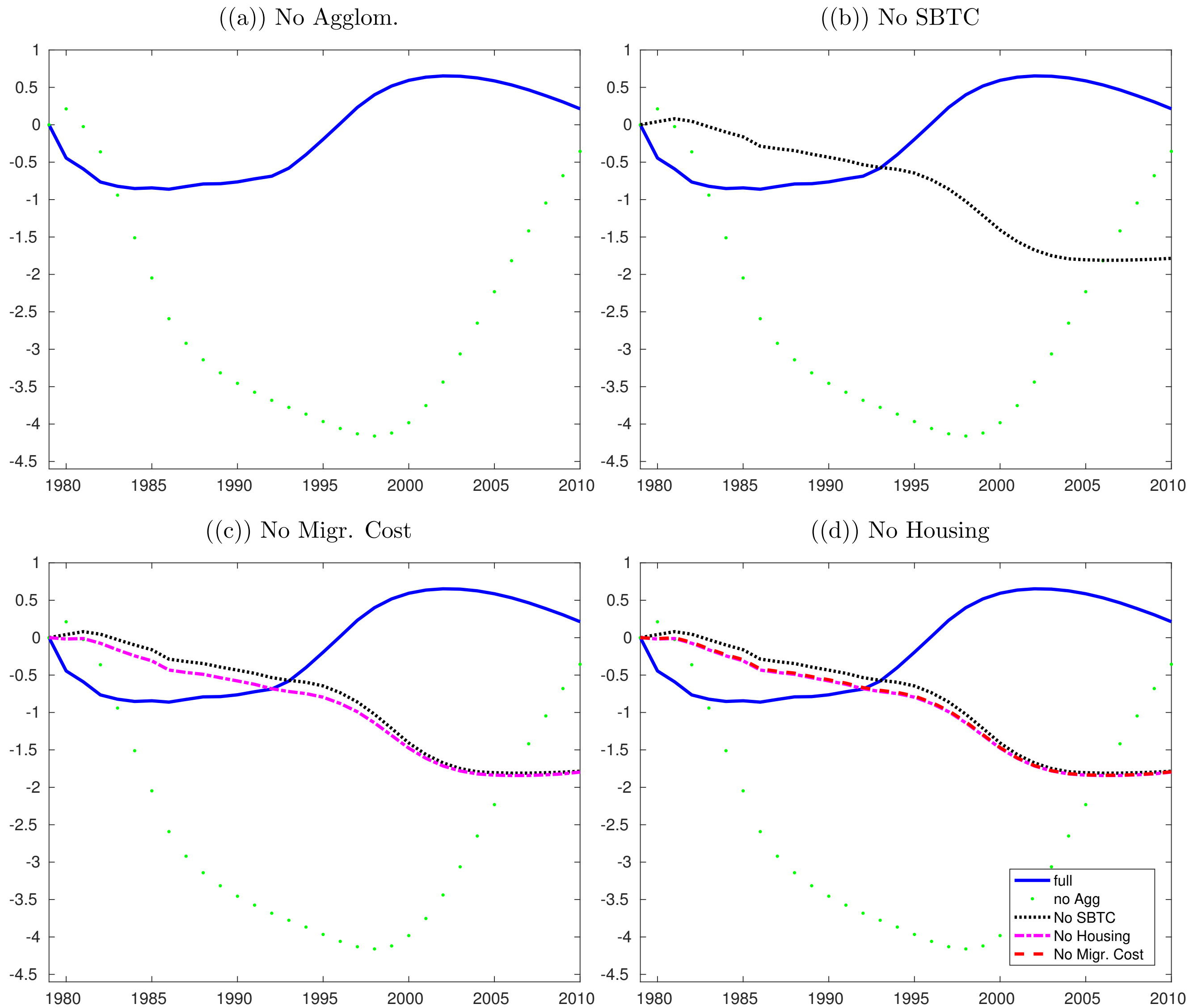
Using model generated data, I estimate specification 4 and compare the outcomes in the model and in the data in figure 16. The model reproduces shows a β^{skill} convergence rate going from -2.5% in 1970 to -.5% in 2010. Overall, β^{skill} declines much faster despite the last period, where it coincides with the model. Panel A shows the results with a difference between t and τ at 10 years, while Panel B shows the difference is at 20 years. Overall, this finding suggest that the model is consistent with other non-targeted moment, which means that it reproduces not only features of the price data, such as the decline in cross-MSA wage convergence, but also features of the quantity data, such as the divergence in the skill ratio.

Figure 9: Quantitative Decomposition of Wage Convergence



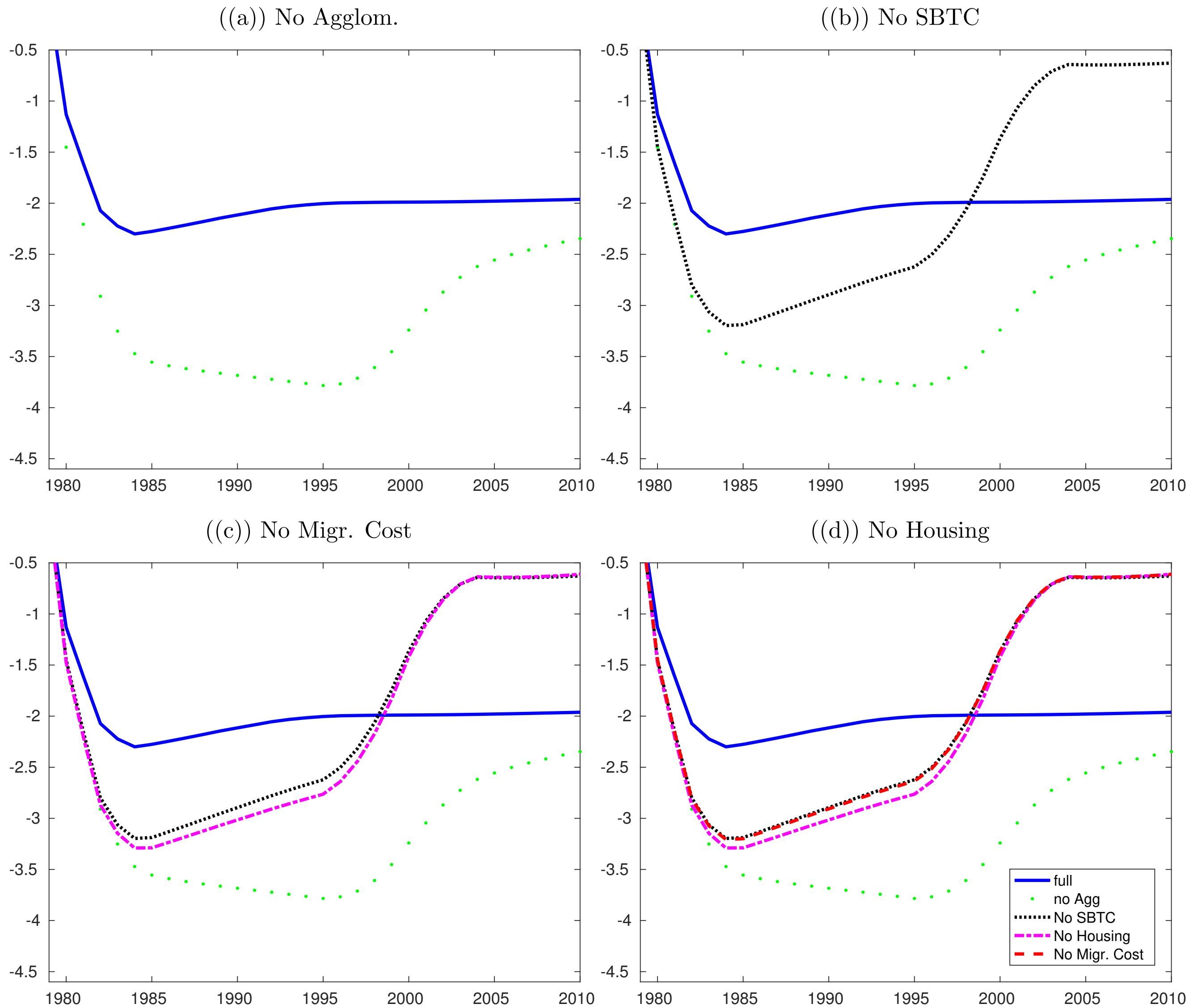
Note: This figure shows the counterfactual exercises in which I turn off the agglomeration forces in subplot (a), SBTC in subplot (b), migration cost in subplot (c), and housing in subplot (d) one after another.

Figure 10: Quantitative Decomposition of Highly Skilled Wage Convergence



Note: This figure shows the counterfactual exercises in which I turn off agglomeration forces in subplot (a), SBTC in subplot (b), migration cost in subplot (c), and finally housing in subplot (d) one after another.

Figure 11: Quantitative Decomposition of Less Skilled Wage Convergence



Note: This figure shows the counterfactual exercises in which I turn off agglomeration forces in subplot (a), SBTC in subplot (b), migration cost in subplot (c) and finally housing in subplot (d) one after another.

5.5 Decline in Gross Migration Flows over time

The decline in geographic migration 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 14 shows that the model matches the data for the trends in the migration rate over time.

5.5.1 Sorting of Highly skilled Migrants to Rich Cities

In section 2, I show that highly skilled workers move more and more to highly skilled MSAs. In this subsection, I check how the model matches this feature of the data. The model, however, does not distinguish migrants from non-migrants. But, it does calculate the migration rates by taking the differences in the population of a MSA over time. To check how the model matches, I generate data on changes in the population of highly skilled H in MSA j and on average wages of MSA j, then, I run the following regression:

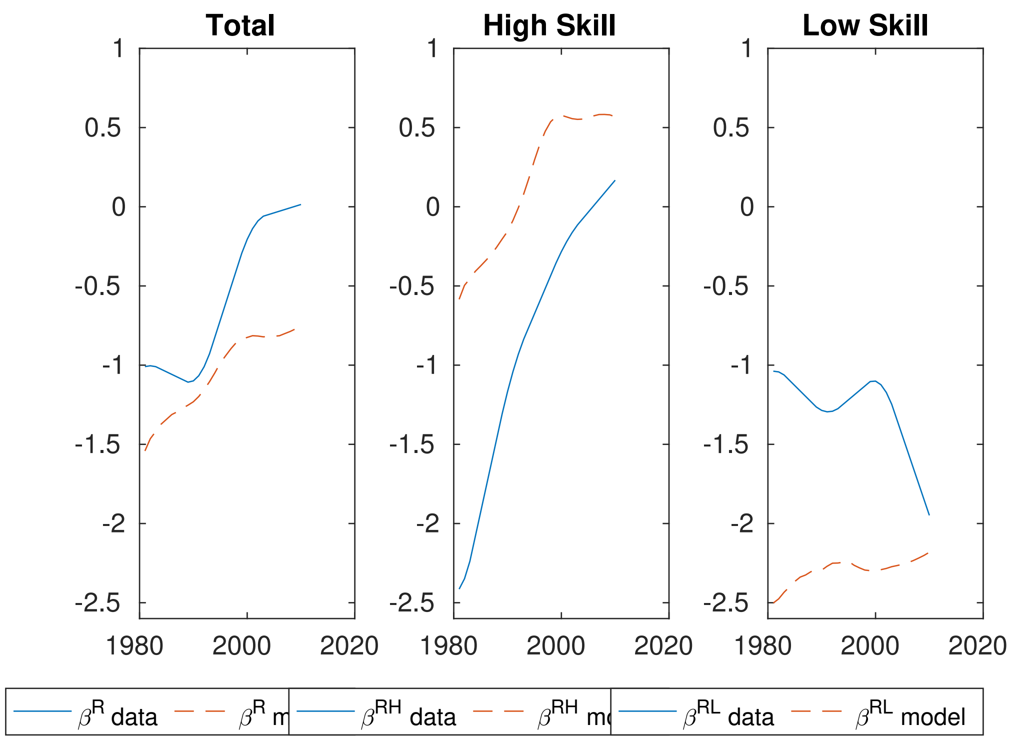
$$\Delta H_{jt} = \alpha + \sum_{t=1941}^{2010} \delta_t^H \ln H_{jt} + \epsilon_t$$

In figure 15, I plot the estimates of δ_t^H for each year. The results show that the MSA with higher average wages, had a decrease in the number of highly skilled migrants between 1940 and 1980. Specifically, a 1% increase in wages generates an approximately 1.4% decrease in the highly skilled migrants. However, the relationship between 1980 and 2010 goes in the opposite direction. Moreover, it increases exponentially over time. In 2010, a 1% increase in local wage increased the number of highly skilled migrants by 2%.

5.6 Real Wages Convergence

With data on wages and local prices generated by the model, I calculate the real wages in each city. In figure 12, I plot the evolution of the β estimate for real wages. I find that the model reproduces a decline in the real wages' convergence as well as in nominal wages. Unfortunately, the literature on real wages historically is quite thin since the availability of data on local prices is very limited. 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 the less skilled workers is still around 2% a year.

Figure 12: Model matching data on Real Wage Convergence



6 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.

6.1 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 also stated 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 18 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.

6.2 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.[24] The results reported in Table 9 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

[24] 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.

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 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.

6.3 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.

6.4 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.

7 Conclusions and Potential Extensions

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 the end of convergence must distinguish between skill groups. Motivated by this observation, I explain the decline in cities' wage convergence by focusing on the role of the interaction between SBTC and agglomeration.

I provide a novel dynamic spatial equilibrium model with heterogeneous agents, local agglomeration spillover, skill-biased productivity shocks, and selective migration. I motivate the assumptions of the models with three novel empirical facts that link together the skill premium, skill concentration, and internal migration: 1) wage convergence declines only among highly skilled workers after 1980, 2) the skill premium is higher in educated cities after 1990, and 3) over time, highly skilled workers begin to migrate to educated cities relatively more than less skilled workers.

I estimate the model with a GMM estimation procedure that uses a SBTC shock and housing regulations as sources of exogenous variation. The model estimates are consistent with the conclusions in the literature. Further, I use model estimates to calibrate some parameters and other parameters from the literature to compute the equilibrium of the model. The calibrated model provides a good fit to the data and shows that both SBTC and agglomeration play an important role in explaining the decline in wage convergence among the highly skilled workers. The main findings show that SBTC explains approximately 80% of the decline in cross-MSA wage convergence in the US after 1980.

Moreover, the model matches other non-targeted moments such as the increase in wage dispersion over the last 40 years that is documented by Hsieh and Moretti (2015), "The Great Divergence" in skills addressed by Moretti (2012), and the secular decline in geographic migration.

This paper is one of the first to study the interaction between agglomeration and SBTC. Moreover, to the best of my knowledge, it is also one of the first to look at the long-run changes in this interaction. 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 decline in poorer MSAs is an important challenge for policy makers. Moreover, understanding regional inequality contributes to understanding the skill premium between highly skilled and less skilled workers. However, the mechanism proposed cannot entirely explain the decline in convergence and the change in the skill premium because of the complexity of the phenomenons, I also think that it captures an important component of them and it might also have external validity when I ask why cross-country convergence does not hold as Barro and i Martin (1992) show. This paper plants the seeds for a broader research agenda. Among others, three open questions are in need of answers. Methodologically, how can we incorporate more features of the data in these large spatial quantitative frameworks. Second, what are the political implications of such increasing geographical differences. Third, 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. Preliminary work indicates the speed of regional convergence has increased across countries but it has decreased within countries.

References

Daron Acemoglu and David Autor. Skills, Tasks, and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 2011a.

Daron Acemoglu and David Autor. *Skills, Tasks and Technologies: Implications for Employment and Earnings*, volume 4 of *Handbook of Labor Economics*, chapter 12, pages 1043–1171. Elsevier, 2011b.

David Albouy. Are big cities bad places to live? estimating quality of life across metropolitan areas. Working Paper 14472, National Bureau of Economic Research, November 2012.

Treb Allen and Costas Arkolakis. Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140, 2014.

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

David H. Autor, David Dorn, and Gordon H. Hanson. Untangling trade and technology: Evidence from local labour markets. *The Economic Journal*, 125(584):621–646, 2015. ISSN 1468-0297. 10.1111/ecoj.12245.

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

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

Robert J. Barro and Xavier Sala-i Martin. Technological diffusion, convergence, and growth. *Journal of Economic Growth*, 2(1):1–26, Mar 1997.

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

Nathaniel Baum-Snow, Matthew Freedman, and Ronni Pavan. Why has urban inequality increased? *American Economic Journal: Applied Economics*, Forthcoming.

William Baumol. Productivity growth, convergence, and welfare: What the long-run data show. *American Economic Review*, 76(5):1072–85, 1986.

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

Kristian Behrens, Gilles Duranton, and Frédéric Robert-Nicoud. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553, 2014. 10.1086/675534.

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

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

Steven Berry, 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*, 112(1):68–105, February 2004.

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

Ariel Burstein, 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.

Lorenzo Caliendo. On the dynamics of the heckscher-ohlin theory. 2011.

Lorenzo Caliendo, Max Dworkin, and Fernando Parro. Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *working paper*.

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

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

Diego A. Comin and Martí Mestieri Ferrer. If technology has arrived everywhere, why has income diverged? *American Economic Journal: Macroeconomics*, (19010), Forthcoming.

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

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

Klaus Desmet, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg. The geography of development: Evaluating migration restrictions and coastal flooding. (21087), April 2015. 10.3386/w21087.

Rebecca Diamond. The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524, March 2016. 10.1257/aer.20131706.

Gilles Duranton and Diego Puga. Micro-foundations of urban agglomeration economies. 4:2063–2117, May 2004.

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

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

Pablo Fajgelbaum and Cecile Gaubert. Optimal spatial policies, geography and sorting. Working Paper 24632, National Bureau of Economic Research, May 2018.

Robert Feenstra 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*, 114(3): 907–940, 1999.

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

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

Edward L. Glaeser and Joseph Gyourko. Housing dynamics. (12787), December 2006. 10.3386/w12787.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118 (3):536–598, 06 2010.

Matthew Higgins, Daniel Levy, and Andrew Young. Growth and convergence across the united states: Evidence from county-level data. *The Review of Economics and Statistics*, 88(4):671–681, 2006.

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

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

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

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

Paul Krugman. Increasing returns and economic geography. *Journal of Political Economy*, pages 483–499, 1991.

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

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

Ahlfeldt Gabriel M., Redding Stephen J., Sturm Daniel M., and Wolf Nikolaus. The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189, 2015.

D. McFadden. Conditional logit analysis of qualitative choice behaviour. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press New York, New York, NY, USA, 1973.

Enrico Moretti. Human capital externalities in cities. *Handbook of Regional and Urban Economics*, 2004.

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

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

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

Stephen Redding and Esteban Rossi-Hansberg. Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58, 2017.

Jennifer Roback. Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6):1257–1278, 1982. ISSN 00223808, 1537534X.

Susan Rosen. Wages-based indexes of urban quality of life. *Current Issues in Urban Economics*, 1979.

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, , and Matthew Sobek. Integrated public use microdata series: Version 6.0 [machine-readable database]. 2015.

Albert Saiz. The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125 (3):1253–1296, 2010.

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

Juan Carlos Suárez Serrato and Owen Zidar. Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review*, 106(9): 2582–2624, September 2016.

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.[25]

Highly and Less skilled Workers I follow the previous work such as Acemoglu and Autor (2011b) 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 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):

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

[25] 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).

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 OLSs. 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 6: Model Estimates for 1940-2010

	(1)	(2)	(3)	(4)	(5)	(6)
ρ	0.531*	0.388***	0.451**	0.558**	0.716*	0.394***
	(0.310)	(0.149)	(0.196)	(0.270)	(0.395)	(0.091)
γ^H	0.616***	0.592***	0.435	0.513*	0.431	0.717***
	(0.231)	(0.199)	(0.292)	(0.266)	(0.349)	(0.103)
γ^L	-0.185	0.003	0.030	-0.023	-0.056	-0.163
	(0.117)	(0.083)	(0.069)	(0.089)	(0.091)	(0.122)
ϕ^H	-0.137	-0.646***	-0.454***	-0.471***	-0.262*	
	(0.088)	(0.180)	(0.152)	(0.165)	(0.158)	
ϕ^L	-0.111**	0.142***	0.156***	0.187***	0.164***	
	(0.047)	(0.040)	(0.040)	(0.043)	(0.041)	
θ	0.503***	0.577***	0.513***	0.490***	0.468***	0.482***
	(0.107)	(0.090)	(0.104)	(0.106)	(0.104)	(0.105)
γ^p	0.679***					
	(0.130)					
γ^{pH}			1.918***	1.933***	1.938***	1.884***
			(0.166)	(0.170)	(0.168)	(0.171)
γ^{pL}			-0.420***	-0.354***	-0.351***	-0.372***
			(0.124)	(0.127)	(0.128)	(0.132)
λ	-0.014	-0.084**	-0.114***	-0.125***		
	(0.062)	(0.036)	(0.037)	(0.039)		
λ^H			0.037		0.285*	-0.063
			(0.032)		(0.147)	(0.100)
λ^L			0.034		-0.168***	0.096
			(0.057)		(0.041)	(0.069)
Observations	823	823	823	823	823	823

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
Subsistance level of Housing: $\bar{O}$	0.25	Ganong and Shoag (2017)
Elasticity of Supply Housing: μ	0.4	Ganong and Shoag (2017)
Share of technology: γ^2	0.99	Desmet et al. (2015)
Migration costs: σ^L and β^L	-.065 and -.861	Notowidigdo (2011) (2013)
Migration costs: σ^H and β^H	-.066 and -1.044	Notowidigdo (2011) (2013)
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: 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 10: Change in β of High and Low Skill Model estimates over Time

	Full	**No Agglom.**	**No SBTC**	**No Housing**	**No Migr. Cost**
$\Delta\beta$	-55.63	-41.15	-41.93	-42.88	-42.99
$\Delta\beta_H$	-106.88	-93.73	-2.94	-5.54	-5.75
$\Delta\beta_L$	-12.94	2.88	-72.88	-73.63	-73.66

Note: 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 each column I vary from the full model to removing step-wise the elements of the model.

Table 11: Change in Variance of High and Low Skill Model over Time

	Full	**No Agglom.**	**No SBTC**	**No Housing**	**No Migr. Cost**
ΔVar(W)	46.23	-73.19	-87.34	-87.96	-88.15
ΔVar(W_H)	519.87	19.03	-58.36	-60.90	-60.58
ΔVar(W_L)	-82.45	-98.24	-95.21	-95.23	-95.56

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.

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.

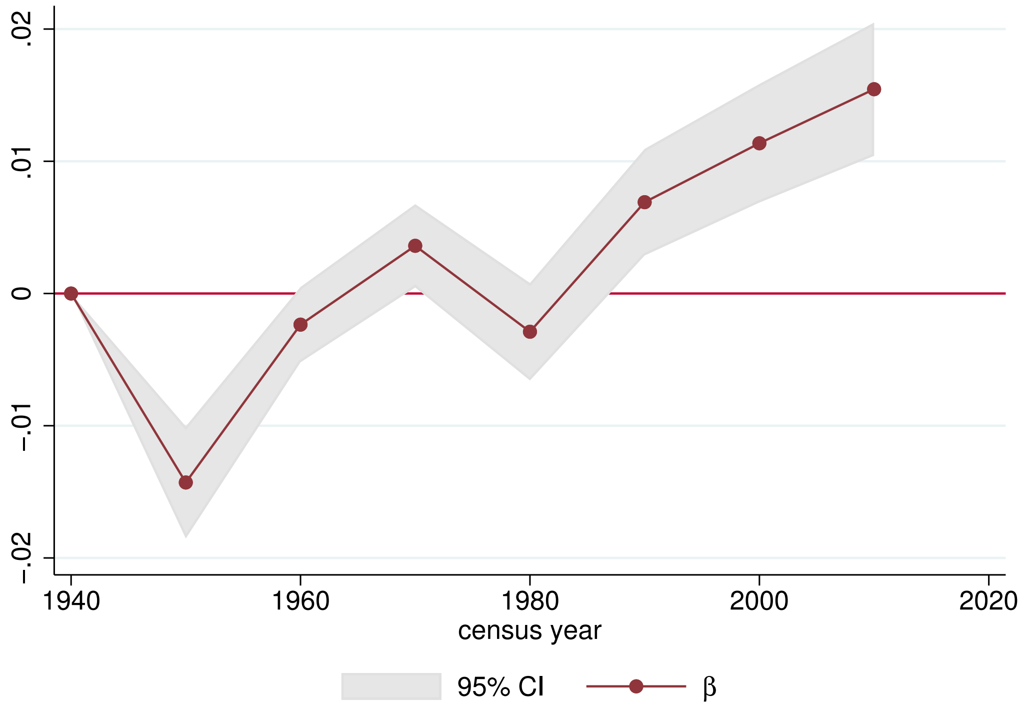
Table 13: Wage Convergence with Import Penetration from China

	(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.

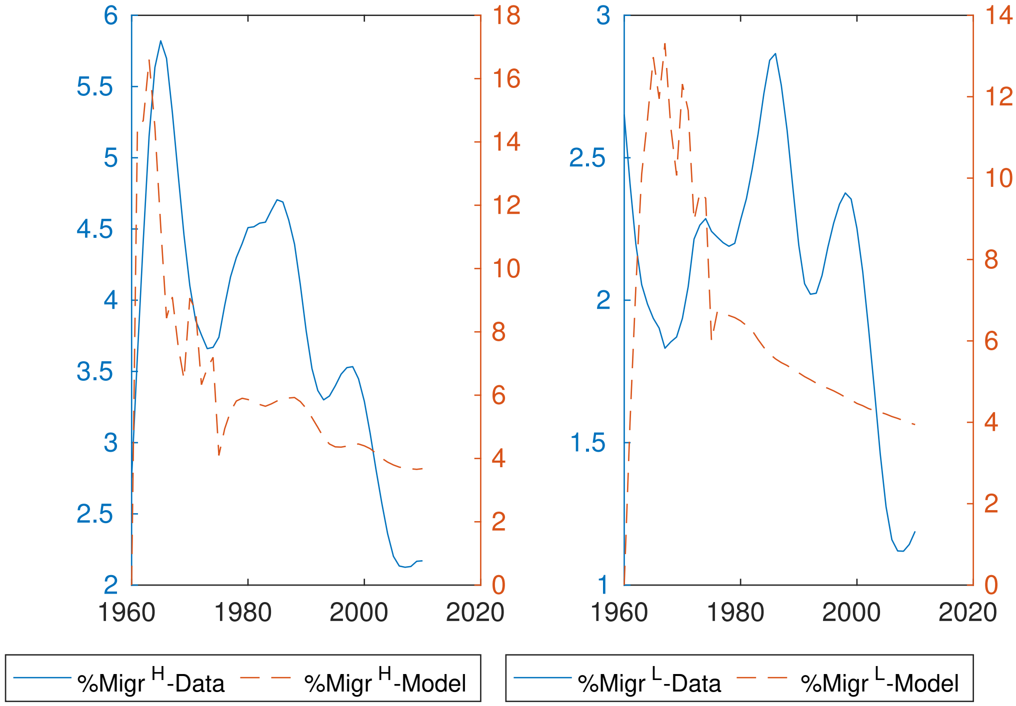
A.3 Figures

Figure 13: 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 14: Migration Rate Over Time

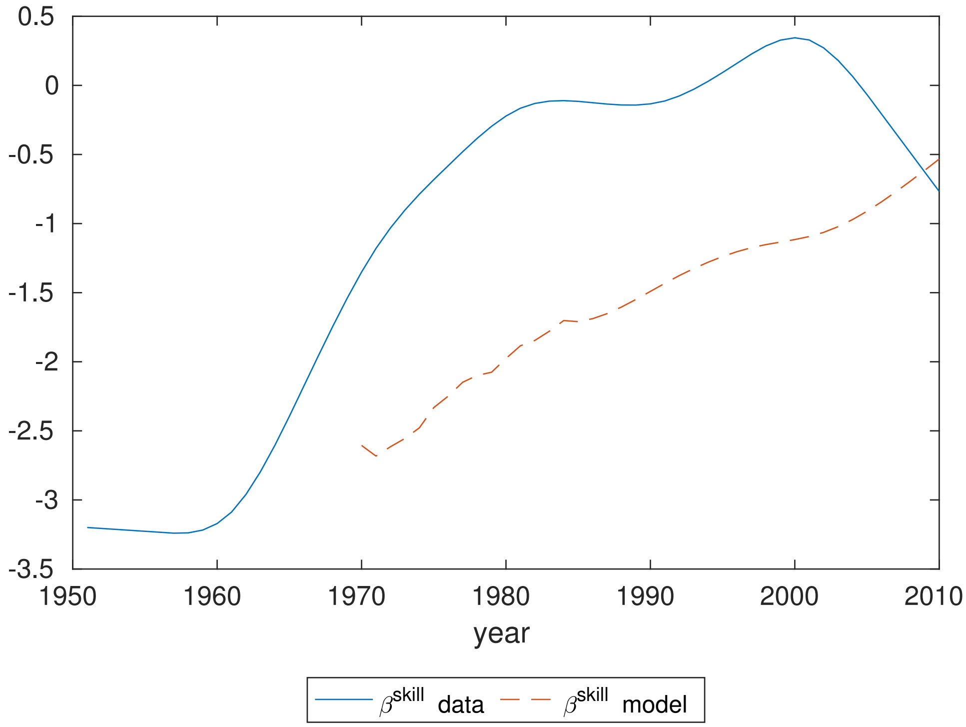


Note: This figure shows the evolution of the migration rate for highly skilled and less skilled workers over time for both the model and the data. On the left, I plot the migration rates generated by the model with a cross and those generated by the data with a circle. On the right plot, I plot the migration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

Figure 15: Sorting Over Time

Note: This figure shows the evolution of the estimates of δ_t^H in equation 5.5.1.

Figure 16: 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.

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

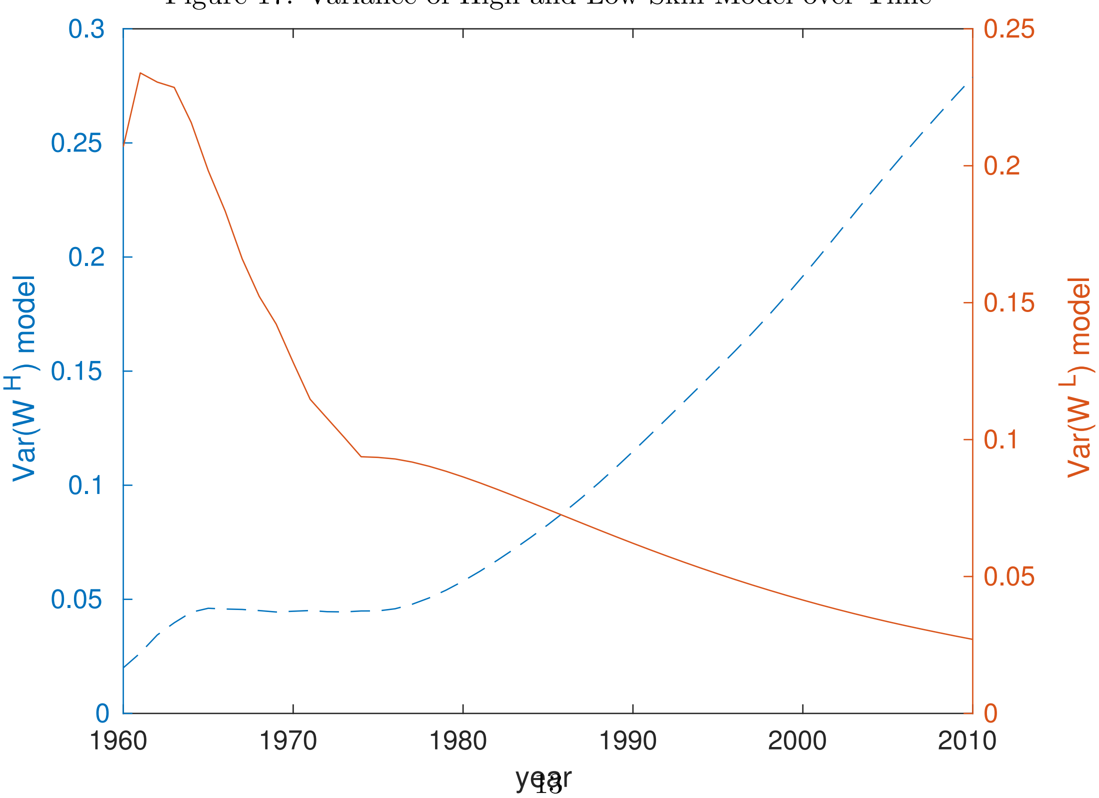
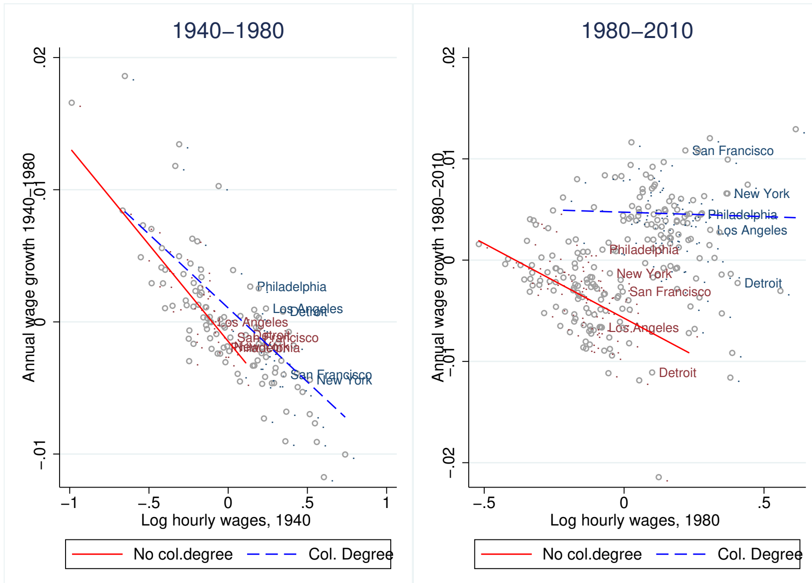
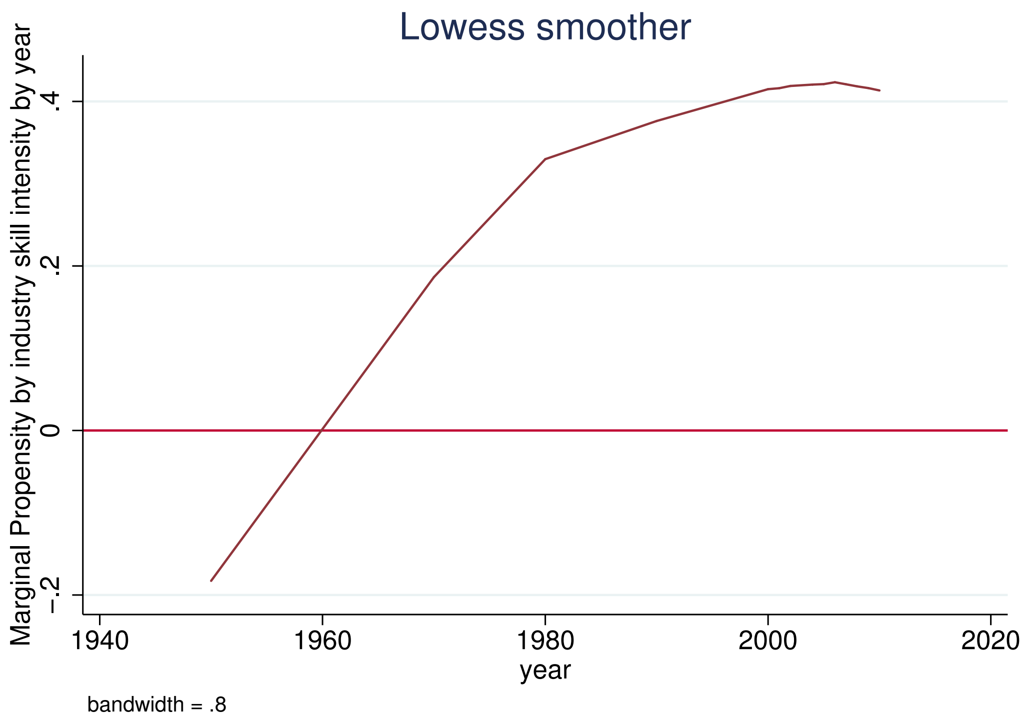


Figure 18: Wage Convergence across MSAs before and after 1980 by Skill Group - Low housing elasticity states



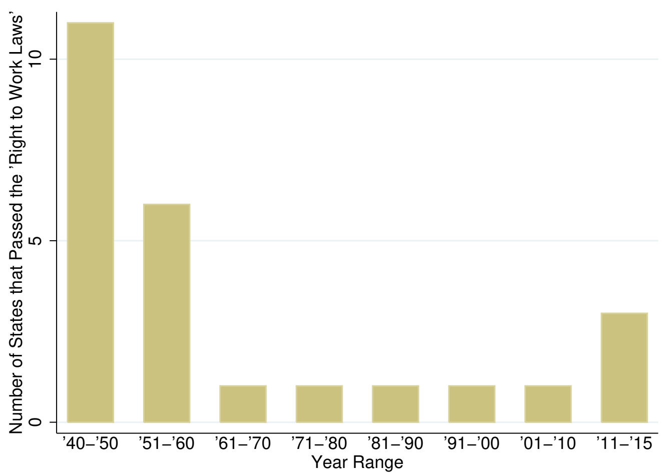
Note: This figure shows two scatter plots of log wages by MSA in the initial year against the annual average growth of wages in the final year by skill type (highly skilled and less skilled workers) in cities that are in states with low housing elasticities. In particular, on the left-hand side (right-hand 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 size of the underlying MSA is represented by the size of the circle in the figure. The line in each graph represents a weighted regression line from the bi-variate regression.

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 6.2.

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.

B Theory Appendix

This appendix supplements the theoretical framework presented in Section 3 in several respects. In subsection B.1, I describe the algorithm for solving the system of equations and obtaining the solution to the model. Subsection B.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 B.3 derives an alternative expression for Y_T.

B.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:

$$W_{Hjt} = (\eta_{Hdjt})[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}H^{\rho-1}_{djt} \tag{16}$$

$$W_{Ljt} = (\eta_{Ldjt})[\eta_{Ldjt}L^{\rho}_{djt} + \eta_{Hdjt}H^{\rho}_{djt}]^{\frac{1}{\rho}-1}L^{\rho-1}_{djt} \tag{17}$$

$$R^{\mu}_{jt} = H_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{18}$$

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

$$H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))} \tag{19}$$

$$L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{20}$$

where

$$\begin{aligned}\delta_{kjt} = \Bigg[&\theta\ln(W_{kjt} - R_{jt}\bar{H}) + \\ &(1-\theta)[\ln((1-\theta)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + (1-\theta)\ln((1-\theta)(W_{kjt} - R_{jt}\bar{O})] + \\ &+A_{kjt} + \gamma^{p}\ln\left(H_{jt}/L_{jt}\right)\Bigg]\end{aligned} \tag{21}$$

and

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

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_{Hjt} and S_{Ljt} are set to zero, then I set the value from the data from S and λ come 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. In particular, the algorithm always converges to equilibrium in a broad neighborhood around the parameter values chosen in the calibration.

B.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. In particular, 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

$$Y_d = \mathrm{K}^{(1-\alpha)}[aH^{\rho} + (1-a)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{22}$$

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

$$Y_d = \mathrm{PC}^{(1-\alpha)}[bH^{\rho} + (1-b)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{23}$$

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

$$\ln\frac{W_H}{W_L} = \begin{cases} \ln\left[\frac{aH^{\rho-1}}{(1-a)L^{\rho-1}}\right] & if\ \frac{H}{L} \leq \phi^L \\ \ln\left[\frac{a\phi^{L\rho-1}}{(1-a)}\right] = \ln\left[\frac{b\phi^{H\rho-1}}{(1-b)}\right] & if\ \phi^L < \frac{H}{L} < \phi^H \\ \ln\left[\frac{bH^{\rho-1}}{(1-b)L^{\rho-1}}\right] & if\ \frac{H}{L} \geq \phi^H \end{cases} \tag{24}$$

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.

B.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 prodictivities 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.

C 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.

C.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 on 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/.[26].

C.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 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.

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

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, the 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. In particular, 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 15 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:

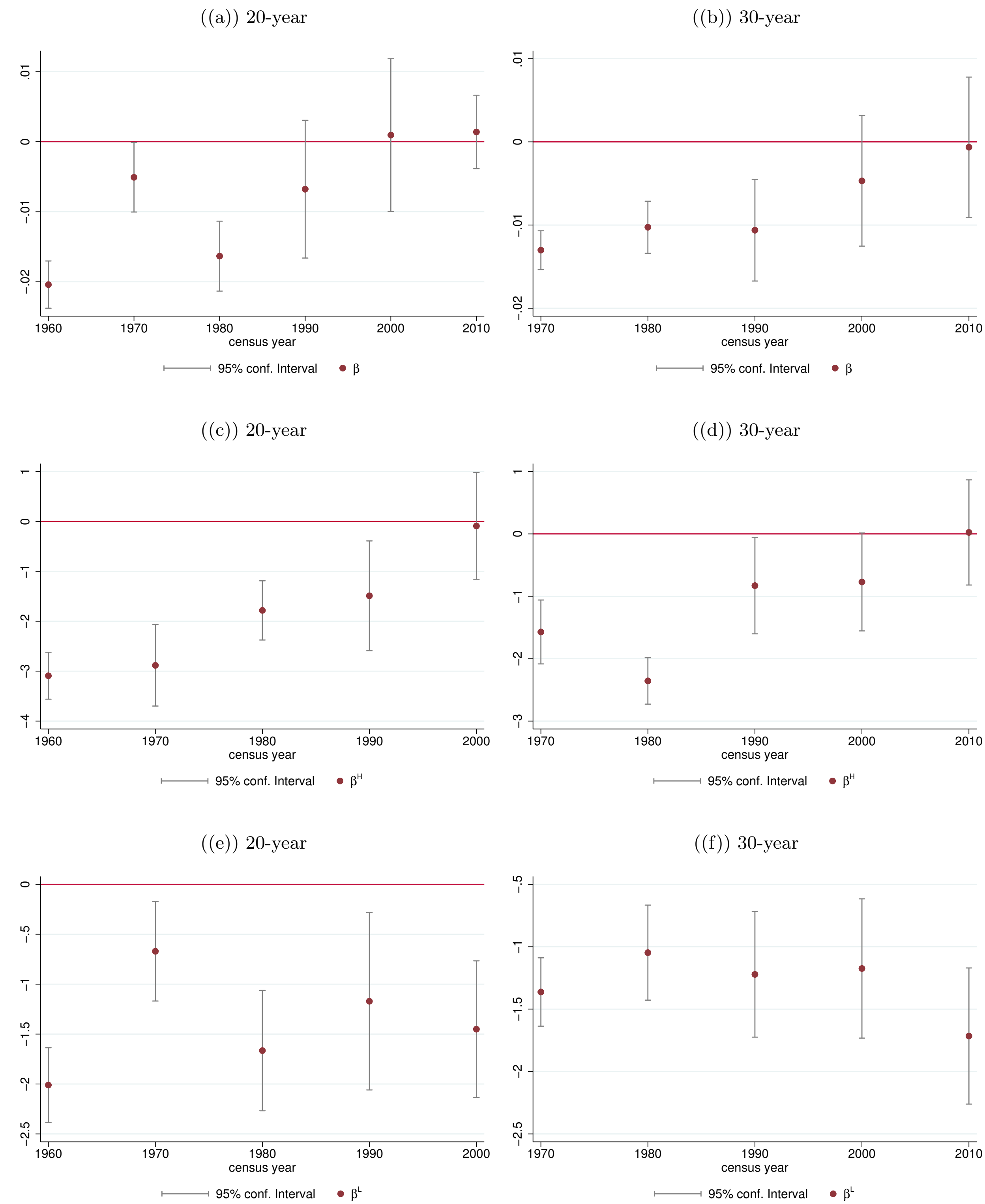
$$\Delta w_{jt} = \beta^o + \beta w_{jt-\tau} + \alpha^H \Delta S_{Hjt} + \alpha^L \Delta S_{Ljt} \tag{25}$$

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

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



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 14: Convergence Rates - Restricted Sample

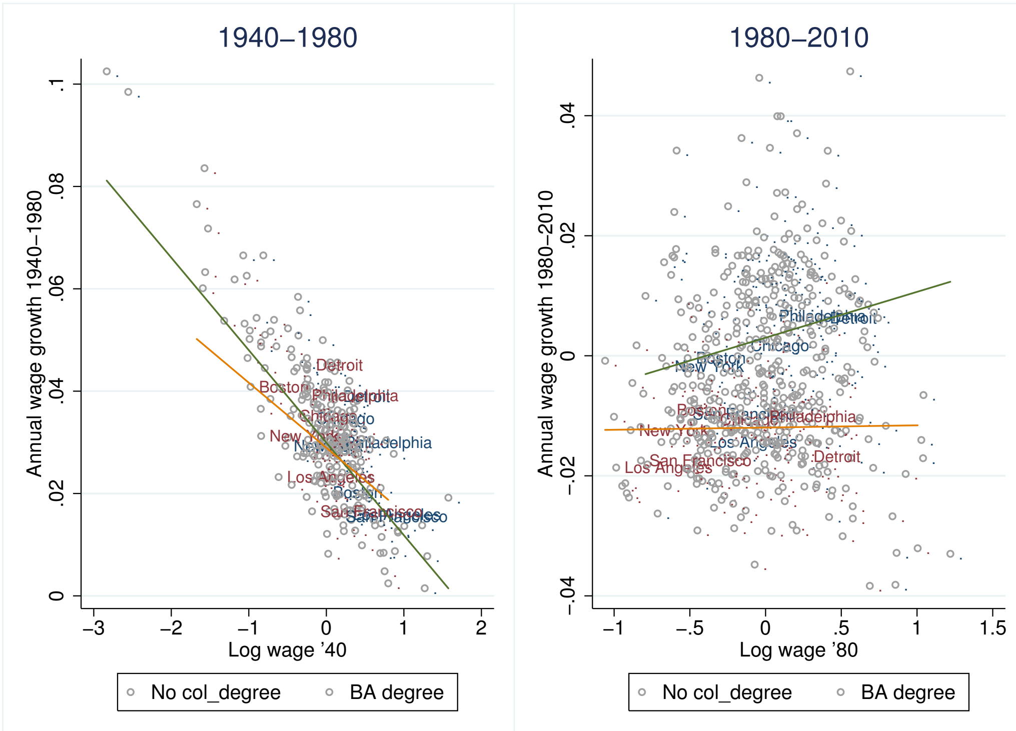
	(1) $\Delta^{1940-1980}$	(2) Δ_{-}^{80-08}
Log($wage^{1940}$)	-0.0109***	
	(-10.53)	
Log($wage^{1980}$)		-0.00116
		(-0.25)
Constant	-0.0217***	-0.0147***
	(-137.22)	(-24.45)

t statistics in parentheses

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

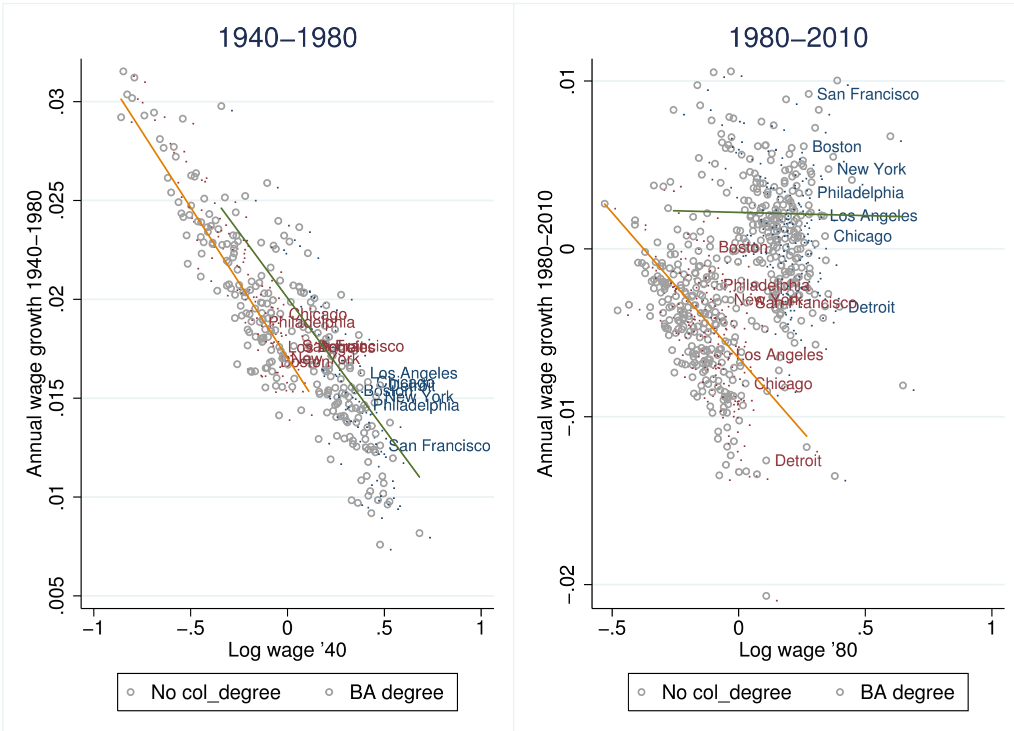
Note: I estimate the β convergence rate for the restricted sample with only 127 cities. In column (1), I estimate it for the 1940-1980 time period and in column (2) for the 1980-2010 time period.

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. In particular, 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 size of the underlying MSA is represented by the size of the circle in the figure. 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. In particular, 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 size of the underlying MSA is represented by the size of the circle in the figure. The line in each graph represents a weighted regression line from the bi-variate regression.

Table 15: 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 un-weighted 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 16: 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.

C.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.

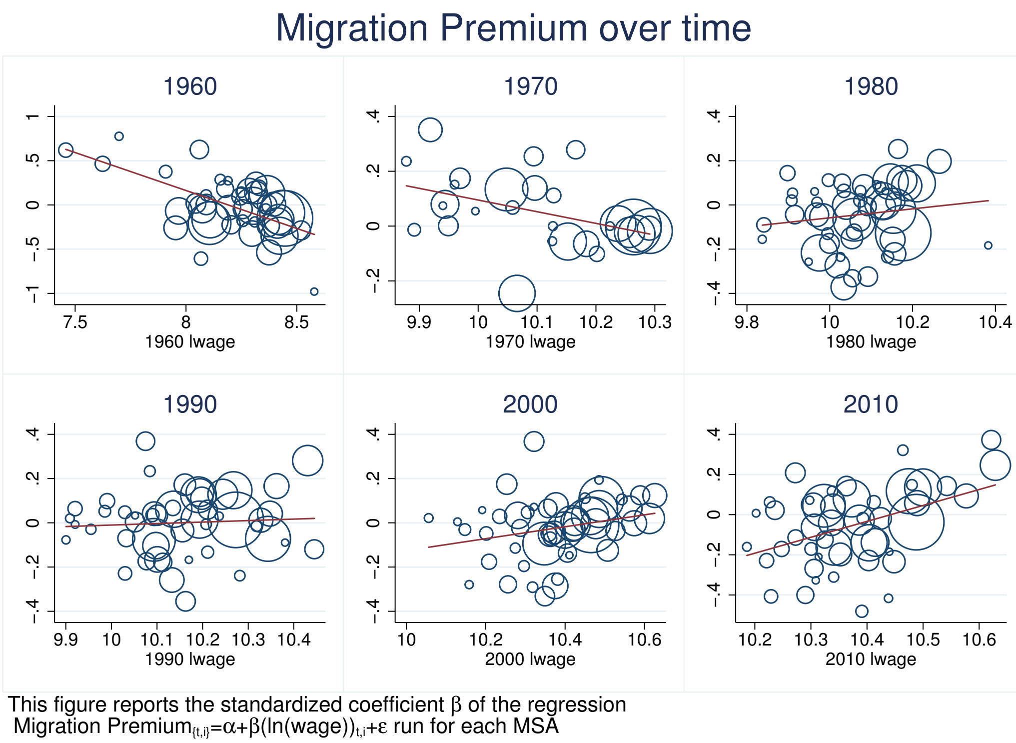
In figure 25, I look at the migration premium over time across states. For each of the years in the CPS sample, I run the following specification:

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

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 states. 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.