# Skilled-Biased Technical Change and Regional Convergence*

Elisa Giannone

January 4, 2017

*Click here for the latest version*

**Abstract**

Between 1940 and 1980 the wage gap between poorer U.S. cities and richer ones was shrinking at an annual rate of roughly 1.4%. After 1980, however, there was no further regional convergence overall. This paper quantifies the contributions of skill-biased technical change (SBTC) and agglomeration economies to the end of cross-city wage convergence within the U.S. between 1980 and 2010. I develop and estimate a dynamic spatial equilibrium model that looks at the causes of regional convergence and divergence. The model choice is motivated by 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 U.S. regional wage convergence. Moreover, the model also reproduces changes in the skill ratio across U.S. cities, as well as, migration patterns after 1980. Finally, the counterfactual analysis suggests that SBTC explains much of 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 would also like to thank Big Banternghansa, Gady Barlevi, Robert Barsky, Paco Buera, Lorenzo Caliendo, Jeff Campbell, Cristian Dagnino, David Dorn, Jason Faberman, Alejandro Justiniano, Munseob Lee, Kerstin Holzheu, Greg Kaplan, Nuno Paixao and other seminar participants at the University of Chicago International Trade, Capital Theory, Applied Macroeconomics working groups, Chicago Fed brown bag seminars, attendees at the AERUS 2015 at Purdue University, WUSTL 2015, LSE SERC Annual Conference 2016, Marco Fanno Alumni Meeting 2016, NASM 2016 conference, Application Workshop, CWIE workshop, Junior Scholar Conference. 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. Contact information: elisagiannone@uchicago.edu

# 1 Introduction

From 1940 to 1980 the college wage premium fell, but it rose after 1980 (Goldin and Margo 1992). Katz and Murphy (1992) show that both the relative price and relative supply of skill increased since 1980, suggesting an increase in relative demand for educated workers. A large literature named this demand shift Skill-Biased Technical Change (SBTC). In turn, SBTC led to rise in earning inequality (e.g., Card and DiNardo 2002, Levy and Murnane (1992), Bound and Johnson 1992).

National skill premia show similarly timed patterns of regional convergence and divergence of wages. In fact, at regional level, between 1940 and 1980, wages in poorer U.S. 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 $\beta$-convergence rate, is 0.014 between 1940 and 1980, however, between 1980 and 2010, it goes to 0 and it is not statistically significant anymore.[2][3]

This paper links regional wage convergence and national demand for skills. Specifically, I find that wage convergence ceased only among high-skill workers, while wage convergence rates for low-skill workers did not decline at all. This finding is consistent with demand shocks like SBTC that favor high-skill workers.

This paper quantifies the contribution of SBTC to the end of cross-city wage convergence within U.S. in the last 30 years. I develop and estimate a dynamic spatial equilibrium model that analyzes the causes of regional wage convergence and divergence. The model is motivated by novel empirical regularities regarding the evolution of the skill premium and migration patterns by skill over time and across space in the last 70 years. I find that the model successfully matches quantitatively the trends in wages among high-skill workers, as well as the increase in spatial wage dispersion. Moreover, the model also reproduces features of the data on quantities such as the evolution of the skill ratio in the last 70 years. Finally, the counterfactual analysis suggests that SBTC explains the vast majority (approximately 80%) of the decline of regional convergence between 1980 and 2010.

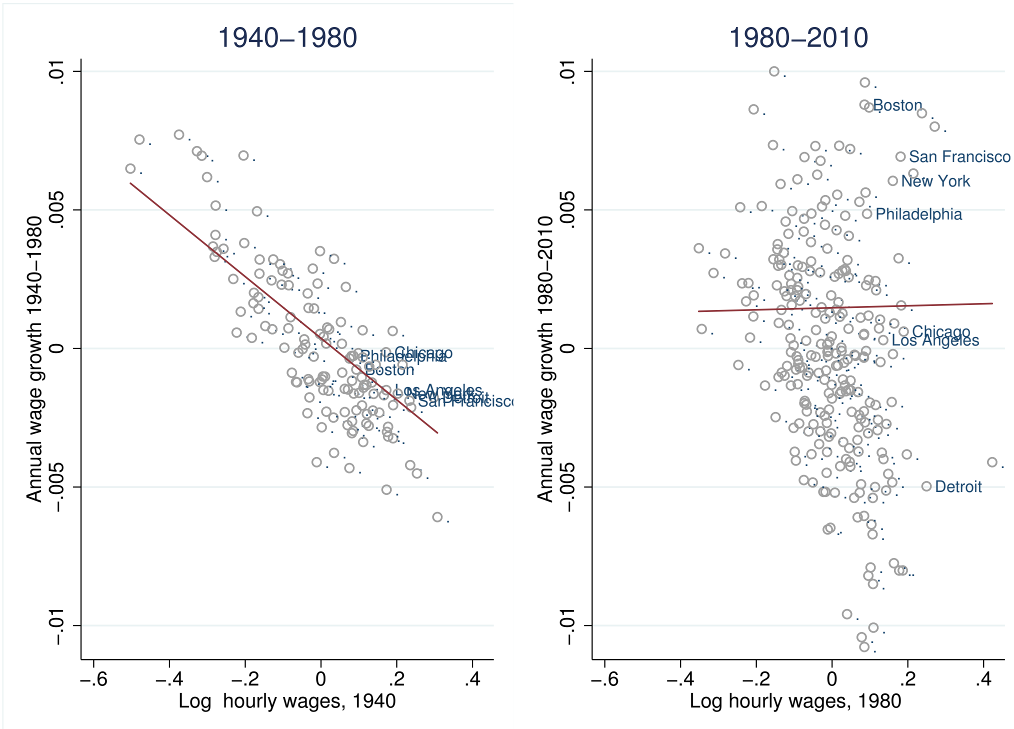
The paper consists of three parts. In the first part, I document a new set of facts about the evolution of wages and quantities by skill across regions during the last 70 years. First, I document that the end of wage convergence occurred only for college-educated workers. Prior to 1980, the wage convergence rate for high and low-skill workers was the same. Since 1980, non-college wages have continued to converge at 1.4% annually, while the rate of wage convergence for college-educated

[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 3.

[2]Berry and Glaeser (2005) are the first to point out to the decline in convergence across cities after 1980. Ganong and Shoag (2015) show a similar decline in convergence for income per capita across U.S. states after 1980.

[3]Data for figure 1 come from U.S. Census and 2010 American Community Survey. A more complete discussion of the sample and wage measures occurs in Section 3.

Figure 1: Wage Convergence across Cities before and after 1980




Note: 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.

wages is 0%. Thus, any account of the end of convergence must distinguish between skill groups. Second, I document that, in the last two decades, the relative price of skill has become positively correlated with relative quantities across cities. Specifically, post 1990, the correlation between college ratio and skill premium is positive across cities. Third, I show that, since 1980, high-skill workers' migration destinations have shifted towards already skill-abundant cities. These facts suggest that performance differences between high-skill (college degree) and low-skill (non-college degree) workers play 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 observations are consistent with SBTC, an important productivity shift that the literature has observed since 1980s. These facts are also consistent with a story of demand forces becoming stronger than supply forces at local level pushing the wages of high-skill 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 literatures. First, in the spirit of Rosen (1979), Roback (1982) and Diamond (2016), I model local labor markets in which workers, high-skill and low-skill, sample the locations where to live that are heterogeneous along wages, rents, local amenities and population in each time 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 document, none of these two sets of models alone can account for regional trends in wages and in employment, differently by agents' skills. These models fail because the spatial equilibrium models lack dynamism and the growth models lack heterogeneity.

The model here allows trends in either direction. Convergence forces enter trough 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 stories, I also introduce other divergence forces such as costly migration and housing.[4]

The model is a dynamic framework in general equilibrium with three sectors, a tradable sector, and two non-tradable sectors (low-skill service and housing). Agglomeration affects only the tradable sector, which is more skill-intensive than the low-skill service sector. The non-tradable service, is a local good so it creates dispersion forces. 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 (Diamond 2016) and a population component. Households decide where to live and when they move, they have a permanent loss in utility. They also decide how much to consume of tradable and non-tradable goods.

The interaction of SBTC and agglomeration economies imply that more educated locations have larger skill premium. High and low-skill workers have some degree of complementarity, so, agglomeration effects raise the wages of all the workers. The differential increase in the wages of high-skill workers makes the migration patterns for high and low-skill workers diverge: high-skill workers migrate to educated cities more than do low-skill workers. Migration has a twofold effect. First, the more workers migrate to a location, the marginal productivity of each will decrease, hence, the returns will decrease. Second, when more high-skill workers move to a location, productivity goes up because of agglomeration effects, raising the wages of all the workers, but especially the wages of the high-skill workers.[5]

Therefore, technological innovation, interacted with agglomeration forces, counterbalances convergence forces that were driving regional convergence until 1980. If technology would not have been skill-biased, convergence forces were favoring the poorer cities, pushing them towards the

[4] Ganong and Shoag (2015) 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 would not account for housing.

[5] The definition of skill premium that I use is the difference between the wages of the high-skill workers compared to the wages of the low-skill workers.

productivity frontier.

In the third part of the paper, I apply the model to the data. For the quantitative application, I build upon Autor and Dorn (2013) to measure SBTC. Autor and Dorn (2013) analyze the effect of computer innovation on regional labor markets output differences. Specifically, when computers arrived, they hit mostly occupations that were very routinized, because machines could replace workers. Therefore, the effect of computers was heterogeneous across locations depending on how many high-skill and low-skill workers that worked in very routinized occupations. Moreover, the degree through which specialization in the local labor market in routine intensive occupations can be predicted by the industry structure 10 years ahead. This motivates the use of city composition of routine task intensive occupations and industries to capture movements on the productivity of the workers and on the labor demand of the firms. The identifying assumption is that SBTC shocks 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, local shocks that capture the skill-biased productivity component interacted with housing elasticities. I estimate the elasticity of population with respect to wages, which has not been conducted in the prior literature. I find that an increase in a city's population raises both high-skill and low-skill wages. Using the estimates from the model, supplemented by others borrowed from previous literature, I calibrate the model and solve it numerically.

The model fits well the data on the patterns of wage convergence and the end of it after 1980 in the last 70 years. In particular, it fits very closely the end of the wage convergence for the high-skill workers. Next, I construct counterfactual exercises by "turning off," step-wise, the divergence forces in the model. The results suggest that the largest share, approximately 80%, of the observed decline in wage convergence among high-skill workers was due to technology becoming more skill biased. Surprisingly, the decomposition of this exercise for high-skill and low-skill workers suggests that if no shock had occurred, convergence would have been, on average, higher for high-skill workers than for low-skill workers.

The model also matches other non-targeted moments very well. Among those, it matches the increase in wage dispersion across cities in the last 30 years. Specifically, Hsieh and Moretti (2015) document a 100% increase in wage dispersion across cities between 1964 and 2009, my model fits this number very closely. Moreover, my model is able to decompose the increase of spatial wage dispersion by skill group. The results suggests that the biggest increase in wage dispersion across cities in the last 40 years happened among high-skill workers and not low-skill workers. Hsieh and Moretti (2015) suggest that the increase in spatial wage dispersion was due to restrictive housing regulations in high productive cities, like New York and San Francisco, in a context where they compared housing and amenities. Through the lens of my model, I can compare how much the increase in spatial wage dispersion was due to housing, migration costs, agglomeration or SBTC. The results suggest 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).

The model, besides matching features of the wages, also matches the skill ratio. Notably, the model suggests that while there was convergence in the skill ratio until 1980, then, afterwards there is divergence, a feature of the data that Moretti (2012) calls "The Great Divergence".

I make several contributions to the literature. To the best of my knowledge, no paper has looked at the joint effects of SBTC and agglomeration effects in a spatial equilibrium setting. Moreover, I disentangle the contribution of SBTC, agglomeration effects, housing and migration on the patterns of wage convergence across cities in the U.S. after 1980. Aside from exploring this new mechanism for regional convergence, the contributions of this paper are also: 1) the exploration of novel facts regarding the differential ending of convergence by skill groups, wage inequality at regional level and migration destination over the last 70 years; 2) the incorporation 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 differentially by skill group and could be used elsewhere.

The remainder of the paper is organized as follows. Section 2 makes a brief summary of the literature to which this paper is connected. Section 3 covers the data and the empirical analysis with description of the empirical facts. Section 4 proposes a theoretical framework. In Section 5, I calibrate the model estimating the core parameters. In Section 6, I solve the model and conduct counterfactual analysis. Section 7 explores other potential complementary channels. Section 8 concludes with a brief summary and future directions.

## 2 Related Work

My work contributes to a number of existing literatures. Most closely related are recent works studying the increase in the spatial dispersion of U.S. cities. Ganong and Shoag (2015) focus on housing supply constraints across states as a mechanism that generated regional income divergence. Their findings are consistent with Hsieh and Moretti (2015) that conclude that housing regulations explain most of the increase in spatial wages dispersion. Hsieh et al. (2013), instead, suggest that the decrease in blacks' discrimination decreased over time reducing regional convergence of income between 1980 and 2008. My paper complements these findings by highlighting the importance of skill-biased technology interacted with local agglomeration forces to explain the decline in wage convergence using as key assumption that productivity before 1980 was not skill-biased but was skill-neutral. To the best of my knowledge, the interaction of the two mechanisms is novel. However, Baum-Snow et al. (2014) discuss the presence of skill-biased agglomeration economies to explain the skill premium across cities.

Besides the patterns in wage convergence, my paper also speaks to the literature of convergence and divergence in skill across cities. Berry and Glaeser (2005) find that human capital across cities is diverging after 1980. Moretti (2012) coins the term "The Great Divergence" of skills, suggesting that high-skill workers are sorting together with other high-skill workers reshaping the job opportunity of U.S. cities. Diamond (2016) looks at skill sorting as another characteristic that may increase the welfare differentials across space. In this paper, I contribute to this literature on skill sorting 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 pre 1980 and the divergence post 1980.

Additionally, this paper speaks to the literature on regional convergence across countries and states, which was inspired by the seminal works of Baumol (1986) and Barro and Sala-I-Martin (1991) and expanded with Barro and Sala-I-Martin (1995). This literature aims at quantifying convergence within and across countries and explaining the reasons for convergence. Bernard and Jones (1996) and Caselli and Coleman (2001) point out to structural transformation as main driver of cross region convergence. In particular, Caselli and Coleman (2001) suggest that since the South has a comparative advantage in farming, at the moment of a structural transformation, workers are pushed out of lower-wage farming and into higher-wage manufacturing. And also, the model foresees higher wages for those that stay in farming. Whereas, Kim (1998) suggests that the economic integration in the U.S. in the second half of the 19th century drove convergence across U.S. states because manufacturing became similar across states. Bernard and Jones (1996) decompose aggregate productivity convergence into industry productivity gains and changing sectoral shares of the output. They find that the manufacturing sector is responsible for the large part of across states convergence in the US. Comin and Ferrer (2013) discuss the decline in convergence across countries despite technology and computers are arriving everywhere. Gennaioli et al. (2014) compare regional convergence across countries and show that barriers to factor mobility explain for convergence differences. My paper complements this literature in several dimensions. First, I provide a realistic model, rather than a North-South model. Second, I propose a model, which had both convergence and divergence forces into it and can match he 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) Beaudry et al. (2010), Autor and Dorn (2013) are more closely related to my paper since besides considering skill-biased technology as a national shock, they focus also on its local implications. In particular, Acemoglu and Autor (2011a) suggest exploiting the difference between the definition of "tasks" and "skills" a principal motivation to capture job polarization when discussing the skill premium.

This paper is also related to the "agglomeration" literature based on Krugman (1991). The "core-periphery" framework offers a global representation of the interaction among scale economies, transport costs and demand and it describes the way these sources generate agglomeration and dispersion of mobile agents through pecuniary externality. Other papers in this agglomeration literature are Behrens et al. (2013) and Greenstone et al. (2010). Duranton and Puga (2003) and Davis and Dingel (2014) propose, instead, a micro-foundation of agglomeration economies of skill.

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 Serrato and Zidar (2014). For example, Diamond (2016) uses Bartik shocks as instruments for wage equations to identify local effects and local amenities. Likewise, Autor and Dorn (2013) use Bartik-type shocks similar to the ones I use in this paper to study routinization share of industry.

# 3 Data and Novel Empirical Regularities

In this section, I briefly discuss the data that I use in the rest of the paper, including the definition of some of the variables that will be used later on. Also, within this section, I report some descriptive facts for the last 70 years. The first three facts are novel empirical evidence that I document. Specifically, I document wage convergence and divergence by skill group, skill premium and migration patterns for cities between 1940 and 2010. The last fact is about the distribution of college and non-college workers across the US. I focus on two skill groups: non-college (those with less than a bachelor's degree) and college (those with a bachelor's degree or more).

## 3.1 Data

Large sample sizes are essential for a detailed geographical analysis of changes in labor market composition. My analysis draws on the Census Integrated Public Use Micro Samples (IPUMS) for the years 1940, 1950, 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 March Current Population Survey (CPS) data that measure migration more frequently and in more detail than the Census data. The CPS is a monthly U.S. household survey conducted jointly by the U.S. Census Bureau and the Bureau of Labor Statistics. The focus is on household and demographic questions, that are asked every month. I use measures of geographic constraints and land use regulations from Saiz (2010). More details about the data can be found in the Appendix.

### 3.1.1 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 MSA by share of high-skill workers over low-skill workers. I define "high-skill" MSAs the ones that have concentration of high-skill workers larger than the national average, the remainder are defined as "low-skill" MSAs. I referred

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

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 high and low-skill, rent by skill group back until 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, the MSAs are not consistent over time. Following the definitions of metropolitan and micropolitan statistical areas, I try to homogenize the definition of MSAs over time. However, this is not possible for all cites.[7]

**High and Low-skill Workers** I follow previous work such as Acemoglu and Autor (2011b) that use education to proxy for skills. Then, I create two groups: "high-skill" workers are the ones who have at least a 4-year bachelor's degree while "low-skill" workers are those who have less years of education 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.[8] 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):

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

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 $X_{it}$. $X_{it}$ includes dummies for age (21-30,31-40,41-50,51-55), one dummy for gender, a U.S. born dummy (whether the worker was U.S. 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.

[7]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 5.1.1 for a definition of commuting zones).

[8]However, for robustness purposes, I checked also the results relaxing the sample restrictions to different extents and, qualitatively, they are unchanged.

**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 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 available for all the years from 1962 to 2009 available in the March CPS, with the exclusion of 1972-1975 and 1977-1979, where information on migration are not collected. 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, college degree (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 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 2.

## 3.2 Descriptive Facts and Novel Empirical Regularities

In this section, I explore novel empirical facts that align with the mechanisms that I want to explain both on wages and on the composition of the skill workers over time and across MSAs. These facts regard the decomposition of wage convergence by skill group, the patterns of skill premium by skill intensity of the MSAs and the migration patterns across MSAs of high-skill workers. First, I document that wage convergence was the same for high and low-skill between 1940 and 1980, but it ended for high-skill between 1980 and 2010. Second, I show that, the correlation between the skill premium and the skill intensity of the MSA is negative until the last two decades but it comes positive afterwards. Third, I show that the probability of migrating is higher for college workers in more educated areas compared to non-college workers and the magnitude increases over time. Fourth, I describe the distribution of high-skill and low-skill across MSAs over the last 70 years.[9]

### 3.2.1 The End of Wage Convergence for High-Skill Workers post 1980

Figure 2 shows that cross-MSAs wage convergence rates between 1940 and 1980 were the same for high-skill and low-skill workers. But, they differ strongly post 1980. Between 1980 and 2010, wage convergence rate occurs only among low-skill workers not for high-skill workers.

Formally, to illustrate these patterns, I run the same "convergence" regression as in Baumol

[9]The fourth fact is partially known to the literature of the "Great Divergence" of Moretti (2012). I describe Moretti (2012) and expand it to a longer time period such as 1940.

(1986):

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

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

Between 1940 and 1980, there was no difference between cross-MSAs wage convergence rates, $\beta_H$ and $\beta_L$. Between 1980 and 2010, instead, the convergence rate $\beta_L$ was still negative and statistically significant, but $\beta_H$ is not anymore. This suggests that the end in convergence was driven only by the wages of high-skill workers, instead, wages of low-skill workers still converge across MSAs. In Panel B of table 1, I report the estimates of $\beta_L$ and $\beta_H$ in the two different time periods both for population weighted and non-population weighted regressions. For population weighted regression, $\beta_L$ and $\beta_H$ are, respectively, -.0123 and -.0143 between 1940 and 1980. Both estimates are highly statistically significant. However, the estimates of $\beta_L$ and $\beta_H$ between 1980 and 2010 are respectively, -.0169 and .000636. The estimate of $\beta_L$ is highly statistically significant but the estimate of $\beta_H$ is not statistically different from 0. In the Appendix, I run several robustness tests for this fact. First, I estimate the rolling convergence for the high-skill and the low-skill, separately for 10 and 20 years windows. Second, I run the same regression as above for compositionally adjusted wages as defined above.[10][11]

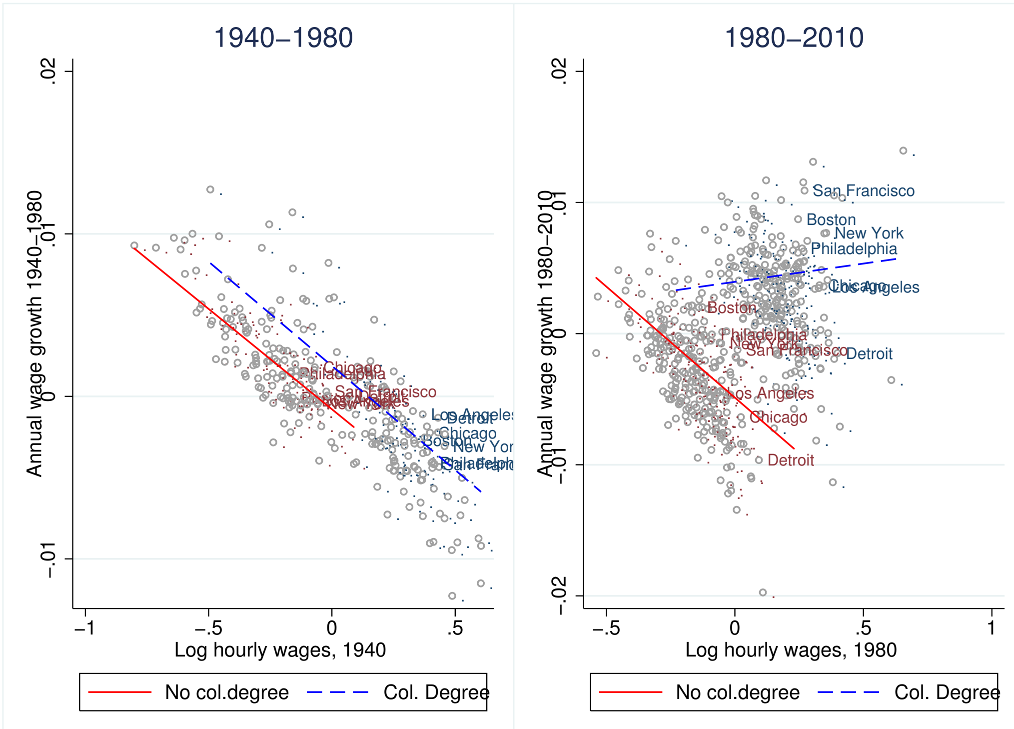
### 3.2.2 Skill Premium

In this subsection, the goal is to show how cross-MSA relationship between the skill premium and the skill ratio has changed over time. It used to be that the skill premium was lower in skill abundant places and in recent years the skill premium is higher is skill abundant places. Figure 3 shows that the skill premium is higher in more educated MSAs conditional on fixed effects post 1990. I define skill-premium as the difference between the wages of the workers with a college

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

[11] Berry and Glaeser (2005) control for changes in college share in the wage convergence relationship.

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 (college degree and less than college degree 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 low and high skill workers. The size of the underlying MSA is represented by the size of the circle in the figure.

degree and the workers without college degree. I run the following regression where:

$$\ln\left(\frac{\hat{w}_{Hjt}}{\hat{w}_{Ljt}}\right) = \sum_{t=1940}^{2010} \beta_t \ln\left(\frac{H_{jt}}{L_{jt}}\right) + \phi_j + \phi_t + \epsilon_{jt} \tag{1}$$

where $\hat{w}_{Hjt}$ and $\hat{w}_{Ljt}$ are the compositionally adjusted wages for MSA $j$ at time $t$ respectively for high-skill and low-skill workers. $\phi_j$ are MSA fixed effects and $\phi_t$ are time fixed effects. $\frac{H_{jt}}{L_{jt}}$ is the ratio of the total number of workers with a college degree and the total number of workers with less than a college degree in MSA $j$ at time $t$. I run the regression for $t = \{1940, 1950, 1970, 1980, 1990, 2000, 2010\}$.[12] Once I run the regression for each year of the Census, I plot the estimate for the coefficient $\beta_t$ for each year. This coefficient can be interpreted

[12]Unfortunately, the Census data for 1960 do not have information on the metropolitan area. Therefore, I need to skip that year.

as an increase in 1 standard deviation of $\frac{H_{jt}}{L_{jt}}$ that is going to affect the skill premium by $\beta_t$ standard deviations. Despite one missing observation in 1960, in figure 3 there is clearly a pattern for the growth of the skill premium by MSA education. In table 2 I report the estimates of $\beta_t$ also controlling for population. In table 3, I report the estimates for annual data using the CPS dataset as a robustness test. Baum-Snow and Pavan (2013) find at least 23% of the overall increase in the variance of log hourly wages in the United States 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. I run the same analysis using CPS data for robustness in figure 12. However, I run the analysis by state, not MSAs, with CPS data. The reason is that CPS data are not very rich in information about MSAs before 1980. The results are confirmed also with this data as shown in figure 12. While, Baum-Snow et al. (2014) find that the growth of the skill premium is negatively correlated with the change in skill ratio in the pool 1980-2007. Complementary to this finding, my analysis in level suggests that the correlation is positive and statistically significant in the 2010 decade, while no clear pattern was present before. 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.

### 3.2.3 Migration[13]

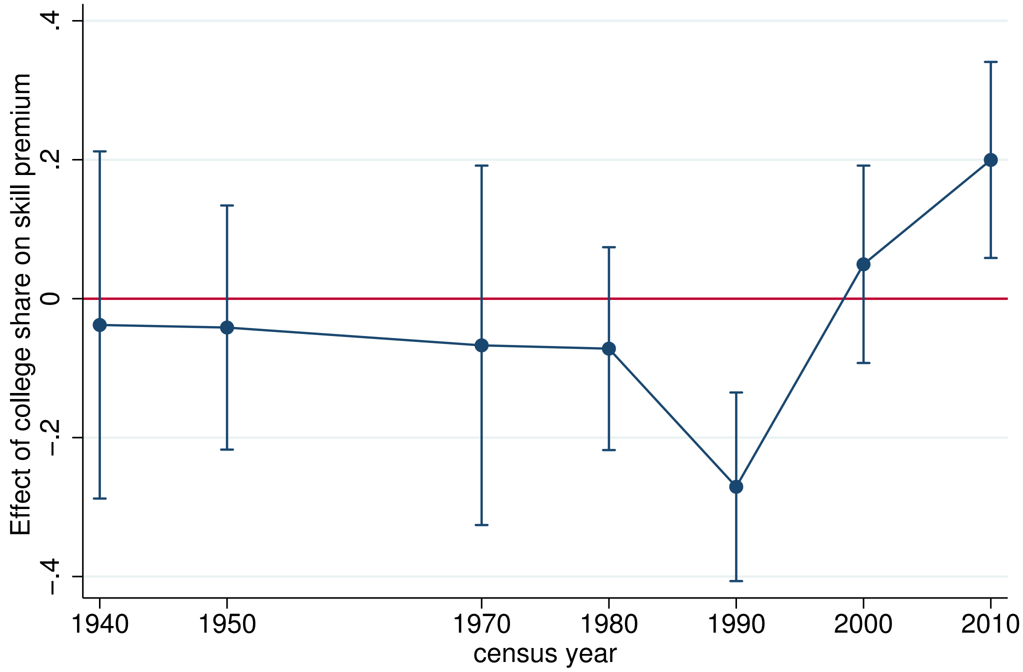
It is well known that educated workers migrate more than less-educated workers. But where are they actually migrating? Are they migrating to less educated places to take advantage of the scarcity of high-skill labor force? In order to assess which type of workers (educated or less-educated) are migrating more to the more educated MSAs, I run a difference-in-difference analysis as in equation (2)

$$1\left(\text{Migrant}_{ijt}\right) = \alpha + \beta 1\left(H_{ijt}\right) + \gamma \frac{H_{jt}}{L_{jt}} + \sum_{t=1963}^{2013} \delta_t 1\left(\mathrm{H}_{ijt}\right) * \left(\frac{H_{jt}}{L_{jt}}\right) + \Gamma X_{ijt} + \phi_j + \phi_t + \mu_{ijt} \quad (2)$$

The dependent variable of this equation is whether worker $i$ in MSA $j$ at time $t$ is a migrant or not. The variable equals 1 if the worker is a migrant. On the right hand side, there is an indicator variable $H_{ijt}$ that equals 1 if the worker is a college degree, 0 otherwise. The second variable is the skill ratio $\frac{H}{L}$ in each MSA and at each time. Third, there is the interaction between the two

[13]The facts reported below all make use of the CPS data. This dataset has better migration information than the IPUMS extracted Census data. Migration information are reported more frequently than for IPUMS extracted Census data. Moreover, the information is more detailed and I can know also whether the migrants migrated across MSAs, not only states as it would be in the IPUMS extracted Census data. 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.

Figure 3: Skill Premium by MSA Education Levels




Note: Figure plots the estimate of the coefficient $\beta$ for the regression 1. On the horizontal axis, I have the decades from 1940 to 2010. While, on the vertical axis, I have estimate of coefficient $\beta$ for each decade from 1940 to 2010. Moreover, there is a line starting at 0 on the vertical axis.

variables I just described. Regression 2 also includes MSA and time fixed effects. The estimated coefficient on this variable $\delta_t$ will be the one I will take into account, using it to compute the marginal effect of being high-skill worker and being in a more skill MSA on the probability of being a migrant. Finally, in the term $X_{ijt}$ there are economic demographics of the workers such as age, gender, race and nationality.[14]

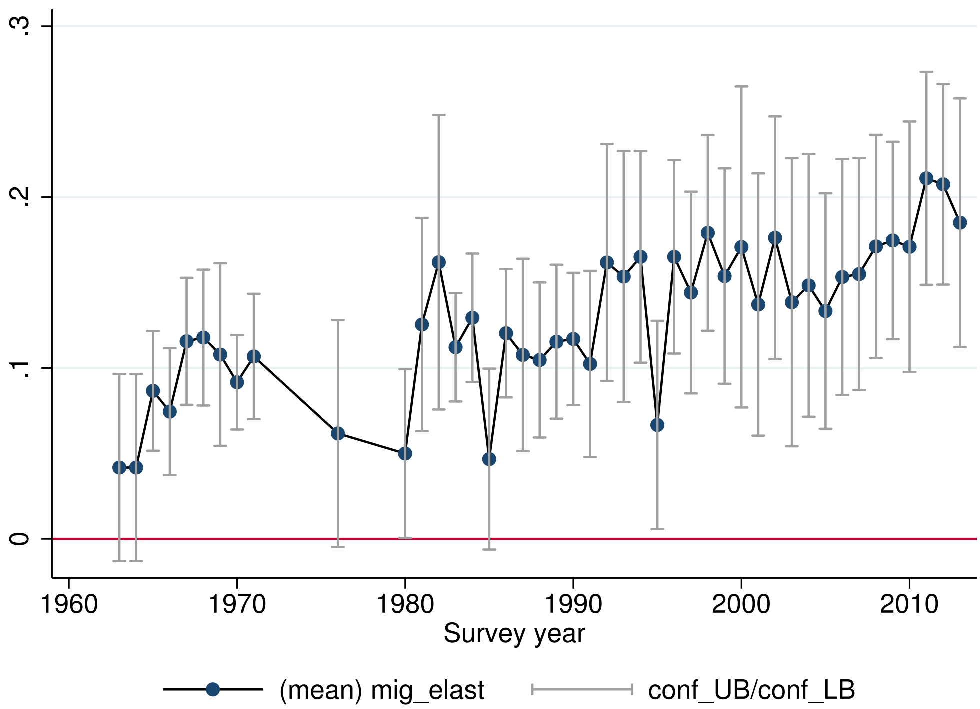
I run regression 2 both as a linear and as a logit model. I focus on the marginal effect of $\delta_t$ to look at the impact of the probability of worker $i$ in MSA $j$ a time $t$ of being a migrant or not given MSA $j$ skill ratio interacted with the worker being a high-skill type. I run the same regression for all the years in the March CPS where the information for about the migration status of the worker was available, which is all the years from 1962-2010, but 1972-1975 and 1976-1979. In the Appendix, I run the same exercise using Census data extracted from IPUMS. Each observation in figure 4 corresponds to the coefficient $\delta_t$ in regression (2). 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,

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

for decade 1960, I take the average of the data available until 1965. For the decade 1970, I take the average of the estimates from 1966 to 1975 and so on and so forth.

Figure 4 shows that the marginal propensity to migrate conditional on being a high-skill workers and moving to high-skill MSAs is increasing over time, in relative terms high-skill workers are concentrating more and more over time in the more educated MSAs. According to these findings, high-skill workers became more likely over time to migrate to higher skill MSAs. This finding goes well in accordance with the hypothesis that high-skill workers are concentrating more and more in educated MSAs. Table 4 shows the evolution over time of the marginal effect of being high-skill and being in a high-skill MSA on being a migrant.

Figure 4: Migration Rate by Destination Education Level




Note: Figure plots the estimate of the coefficient $\delta$ for the regression 2. On the horizontal axis, I have year from 1962 to 2010. While, on the vertical axis, I have estimate of coefficient $\delta$ for each year from 1962 to 2010. Moreover, there is a line starting at 0 on the vertical axis.

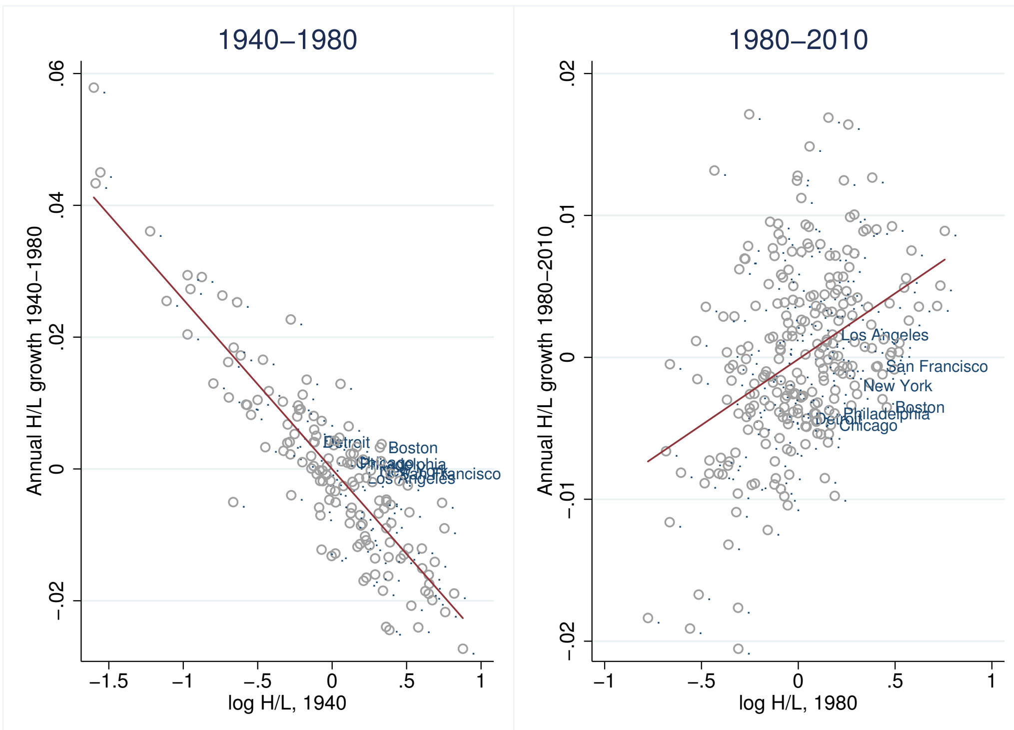
#### 3.2.4 Skill Ratio: The Convergence and the "The Great Divergence" after 1980

What happened to the distribution of high and low-skill workers over time across space? Moretti (2003), Berry and Glaeser (2005), Diamond (2016) and Moretti (2012) show that the skill ratio of workers between 1980 and 2010 was diverging across MSAs. Specifically, Diamond (2016) finds that a 1% increase in skill ratio in 1980 is associated with .17% larger increase in the MSA's college employment ratio from 1980 and 2000 per year. Moretti (2012) coins the term "Great Divergence" to stress how the skills are diverging over space. But what happened to the skill ratio when there was wage convergence across US MSAs? 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 existing literature (Berry and Glaeser 2005, Moretti 2012, Diamond 2016). I estimate the following specification:

$$\log\left[\frac{H_{jt}}{L_{jt}} - \frac{H_{j\tau}}{L_{j\tau}}\right]\frac{1}{(t-\tau)} = \alpha + \beta^{skill} \cdot \log\frac{H_{j\tau}}{L_{j\tau}} + \epsilon \tag{3}$$

where $H$ and $L$ are, respectively, the number of high and low-skill living in MSA $j$ at time $t$ and the initial period $\tau$. The dependent variable is the average annual growth of the skill ratio between $\tau$ and $t$. This regression assesses 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. The difference is that in those regressions, we focused on prices (e.g., wages). Equation 3 focuses on quantities (e.g., the skill ratio of individuals). I run this regression over different time periods using the Census and ACS data. In Figure 5, I plot the observations at MSA level and then the line fit, where $\beta^{skill}$-convergence rates are the slope of the lines. Each circle is an observation by MSA. I label the 10 biggest U.S. MSAs. Between 1940 and 1980, $\beta^{skill}$-convergence rate was negative and statistically significant. However, as suggested by the previous literature, between 1980 and 2010, the $\beta^{skill}$-convergence rate is positive and statistically significant implying skill divergence. Table 5 reports the results decomposing the years in shorter time periods. The results suggest that the distribution of high-skill and low-skill workers across MSAs was converging between 1940 and 1980 and then, it started to diverge between 1980 and 2010. Specifically, In Panel A, I report the results with difference between $t$ and $\tau$ of 10 years. While in Panel B, the same difference is set to be 20 years. As we can see in Panel A, the estimated coefficients are negative and statistically significant until 1970, they become not significant for 1970-1980 and 1980-1990, then, 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 .07% and .04%, respectively between 1990-2000 and 2000-2010. In Panel B, the results are quite similar, however, in column (1), the coefficient is positive and statistically significant. That coefficient is actually calculated between 1940 and 1970 since data for 1960 is not available. Therefore, it could be the case that in a 30 year time span, the results would be 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 college ratio between 1980 and 1980 by .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.

In this section, I showed novel empirical regularities concerning the differential evolution of wages and skills across MSAs in the last 70 years. The first two regard wages. The last two regard migrants and skills. First, wage convergence ended only for high skill workers after 1980, but it was the same for low and high skill between 1940 and 1980. Second, in the last 2 decades, the correlation between skill concentration and skill premium became positive. Third, I showed that high skill migrants are more and more concentrating in high skill MSAs compared to low skill migrants. Fourth, I extend the "Great Divergence" finding of Moretti (2012) for a larger time period such as the last 70 years. These findings provide the motivation for exploring a dynamic spatial equilibrium model with heterogeneous skills.

# 4 A dynamic spatial equilibrium model with heterogeneous skills

The empirical analysis above suggests that skill composition may explain the patterns observed in the skill premium and cross-MSA migration patterns. It also emphasizes the way in which the role of skill composition changed over time. These observations also imply that agglomeration and SBTC effects may be relevant to explain why spatial wage convergence decreased. But how can we disentangle these effects? How important is each of these mechanisms? To attempt to answer these questions, I build a dynamic model of cities. I build on the current spatial equilibrium literature, pioneered by Rosen (1979) and Roback (1982), nested with the literature on economic growth. More specifically, this framework nests several features from Diamond (2016) and Desmet et al. (2015). It departs from Diamond (2016) in several aspects. First, unlike Diamond (2016), I add a dynamic component resembling Desmet et al. (2015). Second, I add a low-skill service sector similar to Autor and Dorn (2013). Third, I emphasize the agglomeration effects in the production side of the economy. In my model, prices are allowed to change across space for the composite good through changes in the prices in the low-skill service sector. This allows me to obtain a series of prices for the low-skill service sector since 1940.[15] Fourth, I shift my 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. While Diamond (2016) focuses on welfare, this paper asks a long-run macro aggregate question. At the same time, this model borrows its dynamic structure from Desmet et al. (2015). However, it departs from the latter by adding agents who are heterogeneous in their skill.

In the model, there are two types of households, high-skill $H$ and low-skill $L$. In each period, they decide how much to consume and where to live. High-skill, H, and low-skill, L, labor are the only two factors of production. Each worker provides, inelastically, one unit of labor in the location where she lives, for which she is compensated with a wage. Each location produces a tradable

[15] This work contributes to the spatial economics literature, which currently lacks a measure of service prices back in time.

good $T$, a non-tradable low-skill service $N$ and non-tradable housing $O$.[16] While the production of tradable $T$ employs both high- and low-skill labor, the low-skill service sector only employs low-skill labor. The productivity terms are different for the two sectors' production functions. While the production function of the low-skill service only depends on an exogenous component, the productivity term of the tradable good's production function contains both an exogenous and an endogenous component. The endogenous component is a function of the ratio of high-skill workers to the low-skill workers, and population. Moreover, worker productivity is different across locations.

## 4.1 Preferences and agents' choices

In each period, agents derive utility from consuming a tradable good $T$, a non-tradable service $N$ and non-tradable housing $O$ according to Stone-Geary preferences. There is also subsistence-level housing $\bar{O}$. Agents also derive utility from exogenous amenities $A_{kjt}$ and from living in bigger cities, with higher $(H_{jt} + L_{jt})$ to some exponent $\gamma^p$. The period utility of an agent $i$ who resides in $j$ at time $t$ and lived in a series of locations $\bar{j}_=(j_0, ..., j_{t-1})$ in all previous periods is given by

$$u_{ikjt\bar{j}_-} = u_{ikjt} \prod_{s=1}^{t} m_k(j_{s-1}, j_s)^{-1} \tag{4}$$

where $u_{ikjt}$ is the utility for agent $i$ of type $k \in \{H, L\}$ at time $t$, which depends only on the current location $j$ of the agents, $m_k(j_{t-1}, j_t)$ is the migration cost of type $k$ of 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

$$u_{ikjt} = \theta \log(T_{kjt}) + (1-\theta)\nu \log(N_{kjt}) + (1-\theta)(1-\nu)(O_{kjt} - \bar{O}) + A_{kjt} + \gamma^p(H_{jt} + L_{jt}) + \zeta_{ijt} \tag{5}$$

where $\zeta$ is a taste shock distributed according to a Gumbell (or Type I Extreme Value) distribution. Thus,

$$\Pr[\zeta_{ijt}] = e^{-e^{(-\zeta_{ijt})}} \tag{6}$$

I assume that $\zeta_{ijt}$ is i.i.d. across locations, individuals and time. Agents discount the future at rate $\beta$ 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, 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

[16] Following Autor and Dorn (2013), I define low-skill services as a sector where there are only low-skill workers providing low-skill services. This includes restaurant meal; janitorial, gardening, haircutting, security services, home health assistance and so on.

all $j \in S$. This assumption turns out to be extremely useful for the feasibility of the model because it implies that agents' choice of location depends only on current variables and not their location history. Therefore, we can rewrite the agents' problem above 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 given by

$$\begin{aligned} V_{kt}(j,\zeta_i') =& \max_{j'}\left[\frac{V_{ikj't}}{m_k(j,j')} + \beta E\left(\frac{V_{kt+1}(j',\zeta_i'')}{m_k(j,j')}\right)\right] \\ =& \frac{1}{m_{k1}(j)}\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')} + \beta E\left(\frac{V_{kt+1}(j',\zeta_i'')}{m_{k2}(j')}\right)\right] \\ =& \frac{1}{m_{k1}(j)}\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')} + \beta E\left(\max_{j''}\left[\frac{V_{ikj''t+2}}{m_{k2}(j'')} + \beta E\left(\frac{V_{kt+2}(j'',\zeta_i'')}{m_{k2}(j'')}\right)\right]\right)\right] \end{aligned} \tag{7}$$

From the last line of equation 7, it follows that the choice of current location is independent of past and future locations. This implies that the value function can be rewritten, isolating the current component as a static problem. Thus,

$$\max_{j'}\left[\frac{V_{ikj't}}{m_{k2}(j')}\right] \tag{8}$$

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

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

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

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

$$\begin{aligned} V_{ikjt} = \Bigg[&\theta\log(\theta W_{kjt} - R_{jt}\bar{O}) + \\ &(1-\theta)(1-\nu)\log\left((1-\theta)(1-\nu)\frac{W_{kjt}}{R_{jt}} + \bar{O}\right) + (1-\theta)\nu\log\left((1-\theta)\nu\frac{W_{kjt} - R_{jt}\bar{O}}{P_{Njt}}\right) + \\ &+ A_{kjt} + \gamma^p\log\left(H_{jt} + L_{jt}\right) + \zeta_{ijt}\Bigg] \end{aligned} \tag{9}$$

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

Using the properties of the Gumbell distribution, following McFadden (1973), we can derive the number of workers of type $H$ and $L$ living in each location $j$ at time $t$.

$$H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))} \tag{10}$$

$$L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{11}$$

where

$$\begin{aligned} \delta_{kjt} = \Bigg[ & \theta \log(W_{kjt} - R_{jt}\bar{O}) + \\ & (1-\theta)(1-\nu)[\log((1-\theta)(1-\nu)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + (1-\theta)\nu \log((1-\theta)\nu \frac{W_{kjt} - R_{jt}\bar{O}}{P_{Njt}}] + \\ & + A_{kjt} + \gamma^p \log\left(H_{jt} + L_{jt}\right) \Bigg] \end{aligned} \tag{12}$$

## 4.2 Technology

In the next subsection, I describe the production technologies of the tradable sector $T$, the non-tradable sectors $S$ and housing $O$. The tradable good is produced using a CES with high-skill and low-skill labor. I split the non-tradable sector in a local non-tradable service sector and housing. The local non-tradable service sector uses a linear production function in low-skill labor. The housing sector is produced depending on the price of the housing sector mimicking Ganong and Shoag (2015). Because the tradable good $T$ is freely tradable across locations, then, the price of $T$, $P_{Tjt} = p_{Tjt}$ $\forall j$, that means that it will be the same across locations and it is assumed to be a nummeraire.

### 4.2.1 Tradable Sector

The production function in equation 4.2.1 is a CES that uses two types of labor $H_j$ and $L_j$ as imperfect substitute inputs.[17]

$$Y_{Tjt} = [\eta_{Ljt} L_{jt}^{g\rho} + \eta_{Hjt} H_{jt}^{\rho}]^{\frac{1}{\rho}}, \qquad \forall\, j = \{1, ..., N\}$$

$\eta_{Hjt}$ and $\eta_{Ljt}$ denote the productivity of $H$ and $L$, respectively, in location $j$ at time $t$. Productivity is divided into an exogenous and an endogenous component.[18] Departing from the standard

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

[18] Applying a change in variable as in Diamond (2016), I can rewrite $Y_{Tjt}$ can be rewritten as a function of data ($w_{Ljt}$, $w_{Hjt}$, $H_{jt}$, $L_{jt}$) and parameters ($\rho$, $\gamma_L$, $\gamma_H, \gamma$):

$$Y_{Tjt} = \left( \frac{w_{Ljt} h^{\rho-1+\gamma_H} l_{Tjt}^{\rho} L_{jt}^{-\gamma_H} + w_{Hjt} h^{\rho+\gamma_H} l_{jt}^{T\rho-1} L_{jt}^{-\gamma_H}}{w_{Ljt} h^{\rho-1+\gamma_H-\gamma_L} L_{jt}^{\gamma_L-\gamma_H} + w_{Hjt} l_{jt}^{T\rho-1}} \right)^{\frac{1}{\rho}} (L_{Tjt} + H_{jt})^{\gamma}$$

formulation of a CES as in Katz and Murphy (1992), I follow Diamond (2016) in order to make productivity dependent on both endogenous and exogenous components. Endogenous differences in productivity depend on the industry mix in the location. To Diamond (2016) I add an extra endogenous component for population spillover. 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.[19] Adding a knowledge spillover through endogenous productivity deriving from the skill ratio is supported also by my empirical findings, as in 3. These two facts suggest that 1) the higher the skill ratio the higher the wage premium in the location and 2) college-educated 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 may be higher in cities with higher concentration of high-skill workers. Simultaneously, following Davis and Dingel (2014) and Baum-Snow et al. (2014), the spillover effects seem to appear also with respect to population, not just the skill ratio.[20] It follows that the expressions for $\eta_{Hjt}$ and $\eta_{Ljt}$ are

$$\eta_{Hjt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (L_{jt} + H_{jt})^{\gamma} S_{Hjt}^{\lambda^H} \exp(\xi_{Hjt}) \tag{13}$$

$$\eta_{Ljt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (L_{jt} + H_{jt})^{\gamma} S_{Ljt}^{\lambda^L} \exp(\xi_{Ljt}) \tag{14}$$

where $S_{kjt}$ is the exogenous skill-biased technology component for $k \in \{H, L\}$. A description of how I account for this shock follows in section 5.1.1.[21] The exogenous productivity component is $\xi_{kjt}$. $\xi_{kjt}$ at time 0 is given and then evolves according to:

$$\xi_{kjt} = \xi_{kjt-1}^{\gamma_2} \left[ \int_s \omega(j,s) \xi_{kst-1} ds \right]^{1-\gamma_2} \tag{15}$$

where $\omega(j,s)$ is a symmetric measure of distance between location $j$ and location $s$ and $\gamma_2 \in [0,1]$.[22] If $\gamma_2 < 1$, then the productivity in location $j$ is dependent on the productivity of the other locations. This will introduce convergence into the model through spatial knowledge diffusion.

More details are given in the Appendix in section B.3.

[19] 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.

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

[21] 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.

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

The profits $\pi$ of the firm will be given by the following maximization problem

$$\pi_{Tjt} = \max_{l_T,h}[\eta_{Ljt}l_T^{\rho} + \eta_{Hjt}h^{\rho}]^{\frac{1}{\rho}} - W_{Hjt}h - W_{Ljt}l_T \tag{16}$$

where $l_T$ is the amount of low-skill labor used by one firm that produces the tradable good. $h$ is the amount of high-skill labor used by a firm that produces the tradable good. A free entry condition drives profits to 0 since the firms will keep entering until the profits are equal to 0. Therefore, a firm choosing its production in period $t$ knows that its current and future profits are going to be equal to 0. This result is extremely useful in solving the model. It means that the dynamic model will be a repeated static model, which facilitates numerical solution of the model.

Since the labor markets are perfectly competitive, the wage in each location will be equal to the marginal product of labor as shown in equations 17 and 18, which derive the first-order condition of the firms.

$$W_{Hjt} = (\eta_{Hjt})[\eta_{Ljt}L_{Tjt}^{\rho} + \eta_{Hjt}H_{jt}^{\rho}]^{\frac{1}{\rho}-1}H_{jt}^{\rho-1} \tag{17}$$

$$W_{Ljt} = (\eta_{Ljt})[\eta_{Ljt}L_{Tjt}^{\rho} + \eta_{Hjt}H_{jt}^{\rho}]^{\frac{1}{\rho}-1}L_{Tjt}^{\rho-1} \tag{18}$$

#### 4.2.2 Low-skill non-tradable service

This sector, which produces low-skill non-tradable services, only uses low-skill labor $L_N$. As in Autor and Dorn (2013), I assume output is a linear function of labor.

$$Y_{Njt} = \exp(\xi_{Ljt})L_{Njt} \qquad j = \{1, ..., N\} \tag{19}$$

Due to competition in the local markets, the wages will be equal to the marginal product of labor:

$$W_{Ljt} = P_{Njt}\exp(\xi_{Ljt}) \tag{20}$$

where $P_{Njt}$ is the price of the non-tradable service good. Because of the constant returns to scale assumption, the profits for all the firms in this sector will be identical and equal to 0. Therefore, all the firms will solve a static profit maximization problem. While it may appear that a low-skill non-tradable sector is not the core of the model, this sector plays several roles. First, it serves as a congestion force. Considering that in my model there are agglomeration economies, it will be useful to have an extra congestion force when housing is shut down. Second, the model is consistent with the literature on spatial equilibrium, such as Serrato and Zidar (2014), which distinguishes between housing and another low-skill service non-tradable. This allows a quantification of both sectors' output. Third, inclusion of a low-skill non-tradable tightens the model's fit to the data since, as explained below, I use an industry composition identification strategy. Therefore, having more than one industry in the model enhances the consistency between the model and the data, though the mapping is still not one-to-one.

#### 4.2.3 Housing Market

The supply of housing is a convex function of its price. The higher the price of housing the higher the supply.[23]

$$O_{jt} = R_{jt}^{\mu} \tag{21}$$

where the exponent $\mu$ represents the elasticity of housing and $R$ is the rental rate of houses in location $j$ at time $t$. This mimics the housing sector following the models of Diamond (2016), Ganong and Shoag (2015) and Serrato and Zidar (2014). 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:

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

### 4.3 Equilibrium

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

**Definition** The equilibrium consists of a set of allocations $\{L_{Tjt},\, L_{Njt},\, H_{jt}\}_{j=1}^{J}$ and a set of prices $\{P_{Njt}, R_{jt}\}_{j=1}^{J}$ , wages$\{W_{Hjt}, W_{Ljt}\}_{j=1}^{J}$ , such that given $\{\xi_{Lj0}, \xi_{Hj0}\}_{j=1}^{J}$, a set of parameters normalizing $P_{Tjt} = P_{Tt} = 1$ 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 utility satisfying equations 10 and 11;
2. Firms maximize profits such that equations 17, 18 and 20 hold;
3. There is free entry for firms into the tradable sector such that $\pi = 0$;
4. Labor markets clear such that 10 and 11 hold;
5. Housing markets clear such that

$$R_{jt}^{\mu} = H_{jt}\left[\bar{O} + (1-\nu)(1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\nu)(1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{23}$$

[23] 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 with Ganong and Shoag (2015). I run some simulations fluctuating the value of the parameter $\mu$ to very large levels and to small levels to check how the housing would respond.

6. The non-tradable service market clears in every location $\forall j \in \{1, ..., N\}$ at each time period $t$

$$(Y_{Njt})^s = (N_{jt})$$

or, in other words:

$$\xi_{Ljt} L_{Njt} = (1-\theta) \frac{H_{jt} W_{Hjt} + L_{jt} W_{Ljt}}{P_{Njt}}$$

7. Technology evolves according to 15.

### 4.4 Discussion

Introducing this persistent productivity formulation with spatial diffusion helps in generating convergence directly in the model, following Barro and Sala-I-Martin (1995), 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 of Barro and Sala-I-Martin (1995), which suggested 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 TFP grows faster in agriculture, there are declining costs of acquiring human capital and farm goods are a necessity good. These two models introduced convergence through two different mechanisms. Also, Caliendo (2011) and Bajona and Kehoe (2010) show that it is possible to prove convergence in a dynamic Hecksher-Ohlin model. The convergence produced by an idea-diffusion process may be related to a declining cost of human capital or to physical capital mobility, as in the previous literature.

The upside of the model used here is that it extends existing spatial equilibrium models to include a dynamic component. This allows us 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, which incentivizes them to switch cities. In a model with no agglomeration forces, a positive number of workers will find it optimal to switch cities and the model would exhibit 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 flows of workers are always positive. Net flows increase as a fraction of workers move to the relatively more productive sector, which 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 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 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 high-skill 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.

### 4.5 Existence and uniqueness

In this section I discuss sufficient conditions for the existence and uniqueness of regular spatial equilibria. Because of the endogenous productivity channels, it is possible that this model will allow for multiple equilibria. This will happen if the agglomeration forces are strong enough that the workers agglomerate all together in the same locations. To avoid this problem, it is necessary to impose restrictions on the parameters governing the production function such that the agglomeration forces are compensated for by dispersion forces. Allen and Arkolakis (2013) prove the existence and uniqueness of 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. They find that the strength of agglomeration and dispersion externalities are crucial to guarantee the uniqueness and existence of a spatial equilibrium. Unfortunately, the proofs of Allen and Arkolakis (2013) and Desmet et al. (2015) do not apply 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 suggest that the values of the agglomeration parameters for which the model has multiple equilibria are definitely higher than the ones I estimate in section 5.1.

## 5 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 values for initial productivity levels, $\xi_{kj0}$ an $S_{kjt}$ for $k \in \{H, L\}$. After obtaining these parameters, I compute the dynamic equilibrium by simply iterating a system of equations. In order to calibrate the model, I estimate the 9 parameters $\{\nu, \theta, \gamma^p, \gamma^L, \gamma^H, \rho, \gamma, \lambda^H, \lambda^L\}$ 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 time periods produces inaccuracies. Second, in order to conduct quantitative rather than mere qualitative analysis, it is important to disentangle the quantitative importance of each of the model's parameters. In particular, it is important to distinguish the effect of agglomeration forces from effects produced by SBTC. Therefore, an identification procedure is necessary to clarify the individual importance of each parameter. I calibrate the other parameters $\{m_{2H}, m_{2L}, \mu, \bar{O}\,\nu, \gamma_2\}$ to data from the literature.

## 5.1 Estimation of the Model

This section describes how I connect the theory to the data. 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 $\Delta S_{Hj,t}$ and $\Delta S_{Lj,t}$ that act as divergence forces while, on the other hand, the other skill-neutral component $\xi_{Ht}$ and $\xi_{Lt}$ pushes poorer cities to reach the productivity frontier more quickly. An important assumption is the linearity between the technology component $\Delta S_{Hj,t}$ and $\Delta S_{Lj,t}$ and the exogenous productivity $\xi_{Hjt}$ and $\xi_{Ljt}$. In the next section, I provide a measure of $\Delta S_{Hj,t}$ and $\Delta S_{Lj,t}$ and specify how this contributes to the complete estimation of the model. The skill-biased component is constructed under Autor and Dorn (2013) assumptions according to which SBTC has a larger impact in locations that are more routine-occupation intensive. Autor and Dorn (2013) show that routinization intensity of a location is negatively correlated with computer adoption.

### 5.1.1 Skill-Biased Productivity Shock

In this section, I provide a measure of the skill-biased productivity component, $S_{Ljt}$ and $S_{Hjt}$ in the model. Therefore, I construct a shock following Autor and Dorn (2013) to capture the adoption of computers.[24] Autor and Dorn (2013) rank commuting zones by routine occupation intensity.[25] The authors build an index of routinization in which they categorize all occupations by their routinization intensity. Each occupation $\upsilon$ 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, Detroit will be very affected by skill-biased technology (or computerization, in this case) because the share of laborers working in a routine-intensive occupation is very high for both high-skill and the low-skill workers. Using the same approach, I construct the routine intensity for both high- and low-skill workers in each occupation, as shown in equation 24 and 25.

$$\Delta S_{Ljt} = \sum_{\upsilon=1}^{\Upsilon} \left( \frac{L_j}{L_{jt}} - \frac{L_{j-10}}{L_{jt-10}} \right) 1 \left( RTI_{\upsilon} > RTI_{P66} \right) \tag{24}$$

$$\Delta S_{Hjt} = \sum_{\upsilon=1}^{\Upsilon} \left( \frac{H_j}{H_{jt}} - \frac{H_{j-10}}{H_{jt-10}} \right) 1 \left( RTI_{\upsilon} > RTI_{P66} \right) \tag{25}$$

Autor and Dorn (2013) suggest that when the price of computers start falling, then, workers in routinized occupations, who are substitutable by computers, see their wages erode. Therefore, MSAs that specialized in routine occupations, both for high and low skilled workers, experienced

[24] I also compare my results with the results that I would have obtained using the Bartik shock used in Diamond (2016).

[25] 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/

relative wage declines. $\Delta S_{kjt}$ capture well this story through the measure of routinization. Using this same approach, I build the routine intensity in each occupation both for the high-skill and the low-skill workers as in equation 24 and 25. $\Delta S_{Hj,t}$ and $\Delta S_{Lj,t}$ are two good proxies of how SBTC affected cities in different ways depending on their composition.[26] However, this is not a good measure of productivity shock because it correlates with contemporaneous and local changes that could affect wages. Following the approach of Autor and Dorn (2013), I use national employment changes both for the high-skill and low-skill workers interacted with local routine intensity industry share 10 years ago as instruments for $\Delta S_{Lj,t}$ and $\Delta S_{Hj,t}$. These instruments can be described as:

$$\Delta \tilde{S}_{Hj,t-10} = \sum_{ind} \left(H_{ind,-j,t} - H_{ind,-j,t-10}\right)\left(R_{ind,j,t-10}\right)$$

$$\Delta \tilde{S}_{Lj,t-10} = \sum_{ind} \left(L_{ind,-j,t} - L_{ind,-j,t-10}\right)\left(R_{ind,j,t-10}\right)$$

where $-j$ is all cities in the sample other than MSA $j$, $ind$ is industries in the economy and $t$ is time. $H_{ind,-j,t}$ and $L_{ind,-j,t}$ are, respectively, the number of high skill and low skill workers in each industry $ind$ at national level at time $t$ excluding MSA $j$ to avoid mechanical correlations. $H_{ind,-j,t-10}$ and $L_{ind,-j,t-10}$ are the same lagged 10 years. $R_{ind,j,t-10}$ is the routine occupation share among workers in each industry in a specific MSA $j$. Unlike Autor and Dorn (2013), I create both the index and the instrument for high-skill $H$ and low-skill $L$. In this way, I produce extra variation in the data and use the differential impact of technological shocks on the two categories of workers. These instrumental variables, $\Delta \hat{S}_{Lj,t-10}$ and $\Delta \hat{S}_{Lj,t-10}$, are useful in the estimation of model parameters of the model and in the construction of the moment condition.

Table 6 presents first-stage estimates for these instrumental variables. The predictive relationship between $\Delta S_H$ and $\Delta \hat{S}_H$ is sizable and highly significant, with F-stats of 10 or above in each decade as shown in Panel A. The predictive relationship between $\Delta S$ and $\Delta \hat{S}_L$ is sizable and highly significant, with F-stats of 10 or above for the decades after 1980. However, the F-stats for decades 1950, 1970 and 1980 are less than 10. In particular, in 1970, the F-stat is less than 7.[27] The sections below discuss how these labor demand shocks are used in identifying the model parameters. Since the estimation employs changes in wages and population, the changes in labor demand will provide an appropriate source of variation for cities over time. More details are provided below.

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

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

#### 5.1.2 Labor Demand

In order to estimate labor demand I use moment conditions starting from the labor demand curves for high- and low-skill workers. The change in productivity levels interacted 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 $\{\gamma_H, \gamma_L, \gamma, \rho, \lambda^H, \lambda^L\}$.

For this purpose, I start by taking the logs and the first differences of the equations 17, 18 and 4.2.1:

$$\Delta w_{Hjt} = (1-\rho)\Delta \ln Y_{Tjt}(\rho, \gamma_{H,} \gamma_L) + (\rho-1)\Delta \ln H_{jt} + \gamma_H \Delta \ln \frac{H_{jt}}{L_{jt}} + \gamma \Delta \ln (H_{jt} + L_{jt}) + \lambda^H \Delta S_{Hjt} + \Delta \xi_{Hjt} \tag{26}$$

$$\Delta w_{Ljt} = (1-\rho)\Delta \ln Y_{Tjt}(\rho, \gamma_{H,} \gamma_L) + (\rho-1)\Delta \ln L_{Tjt} + \gamma_L \Delta \ln \frac{H_{jt}}{L_{jt}} + \gamma \Delta \ln (H_{jt} + L_{jt}) + \lambda^L \Delta S_{Ljt} + \Delta \xi_{Ljt} \tag{27}$$

I follow the same approach for the service sector from equation 20:

$$\Delta w_{Ljt} = \Delta \ln P_{Njt} + \Delta \xi_{Ljt} (28)$$

As in Diamond (2016) and Serrato and Zidar (2014), the identification strategy follows from changes in the labor supply which are uncorrelated with local productivity. Also, the interaction of SBTC shocks with cities' housing supply elasticities leads to variation in the labor supply that is uncorrelated with the unobserved changes in local productivity. The housing supply affects migration decisions in response to a labor demand shock. Differential housing supply elasticities will give me exogenous variation in labor supply. Compare two cities: one has a very elastic housing supply and the other has a very inelastic one. Both cities 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, offsetting the increase in wage through the labor-demand channel. Specifically, the exclusion restrictions are:

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

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

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

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. In the interest of brevity, I report the details of these moment conditions in the Appendix.

#### 5.1.3 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

$$V_{ikjt} = \delta_{kjt} + \zeta_{ijt} \tag{29}$$

where

$$\begin{aligned} \delta_{kjt} = \Bigg[ & \theta \log(W_{kjt} - R_{jt}\bar{O}) + \\ (1-\theta)(1-\nu)\left[\log((1-\theta)(1-\nu)\frac{W_{kjt}}{R_{jt}} + \bar{O})\right] + (1-\theta)\nu & \left[\log((1-\theta)\nu\frac{W_{kjt} - R_{jt}\bar{O}}{P_{Njt}})\right] + \\ & A_{kjt} + \gamma^{p}\log\left(H_{jt} + L_{jt}\right)\Bigg] \end{aligned} \tag{30}$$

The fact that the model does not rely on agents' history simplifies the estimation procedure by causing it to resemble a static framework. The estimation of the labor supply follows from the decision of the agents of where to live in each period. Because the utility component $\delta_{kjt}$ doesn't depend on individual worker characteristics, the estimates for each type $k$ will be exactly equal to the log population of each demographic group observed living in the MSA. Therefore, this is a simplification with respect to Berry et al. (2004). We take the difference in mean utility $\delta_{kjt}$ over time to get

$$\begin{aligned} \Delta\delta_{kjt} = \theta\Delta\log(W_{kjt} - R_{jt}\bar{O}) + (1-\theta)(1-\nu)\Delta\left[\log((1-\theta)(1-\nu)\frac{W_{kjt}}{R_{jt}} + \bar{O})\right] + \\ (1-\theta)\nu\Delta\left[\log((1-\theta)\nu\frac{W_{kjt} - R_{jt}\bar{O}}{P_{Njt}})\right] + \\ \Delta A_{kjt} + \gamma^{p}\Delta\log\left(H_{jt} + L_{jt}\right) \end{aligned} \tag{31}$$

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 $\Delta A_{kjt}$. This reasoning follows Diamond (2016). Specifically, I use SBTC shocks and their interaction with supply elasticity characteristics. 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 local exogenous amenity changes. 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, $\theta$, and the share of service with respect to housing, $\nu$, from the literature. Beraja et al. (2016) consider a value of $\theta$ of 0.6. and Serrato and Zidar (2014) consider a range of $\theta$ from 0.3 to 0.6. Moreover, Serrato and Zidar (2014) consider a value for the share of the non-housing non-tradable equivalent to $\nu$ equal to 0.4. I will also estimate the model without using the externally calibrated data, relying only on the instruments for identification. In particular, the moment restrictions are:

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

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

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

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. Finally, I test whether the over-identification restrictions can be jointly satisfied.

## 5.2 Migration Costs

By taking differences of $\delta_{kjt}$, migration costs $m_{k2}(j)$ are eliminated since they do not vary over time. Therefore, another strategy is needed to calibrate the migration costs. One potential approach is to calibrate them using the existing literature. In particular, we can use the estimate from Notowidigdo (2011), which provides separate migration costs for high- and low-skill workers. Notowidigdo (2011) uses an exponential function to estimate migration costs. The functional form he estimates is as follows:

$$m_{k2} = \frac{\sigma^k exp(\beta^k x_j) - 1}{\beta^k} \tag{32}$$

where $x_j$ relates to MSA characteristics such as population. This functional form is very flexible since, despite having only 2 parameters, it has advantageous curvature features.

### 5.2.1 Estimation Results

I estimate the full model described in the previous section. I use a GMM estimation procedure with data at the MSA level for the entire time period, 1940-2010 with data every 10 years except for 1960. The results are reported in table 7. Some of the parameters can be compared with those in the prior literature, and overall I find that their magnitudes to be broadly similar. The results of the estimates of the model are thus in accordance with the existing literature that estimated similar parameters. In particular, the results related to endogenous spillover $\gamma^H$ and $\gamma^L$ are similar to Diamond (2016) which uses similar functional forms. The estimates suggest that returns to education are strong. In particular, I find that a 1% increase in the share of college-educated workers will raise the high-skill wage by .25% and the low-skill wage by .19%. An increase in a MSA's low-skill worker population, on the other hand, does not impact wages for either group. These estimates do differ slightly from Diamond (2016), which finds that they do differ. Similar estimates were also conducted by Moretti (2003), which finds that a 1% increase in MSA's college employment ratio leads to a .16% increase in the wages of high-skill workers and a .16% increase in the wages of low-skill workers. A novel parameter that has not been estimated before is $\gamma$, the population endogenous productivity effect. The estimates suggest that a 1% increase in the population of MSA $j$ will increase high-skill wages by 2.6% and low-skill wages by 2.1%. Baum-Snow and Pavan (2013) estimate that at least 23% of the overall increase in the variance of log hourly wages in the United States 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 go well in accordance with their findings.

While I do not report it in this estimate, I also analyze another specification in which the low-skill service sector has decreasing returns to scale with coefficient $\alpha$. The estimation suggests that this coefficient is very close to 1, meaning that the low-skill service sector has nearly constant returns to scale. Moreover, the elasticity of substitution between high-skill and low-skill workers, which is equivalent to $\frac{1}{1-\rho}$, equals 1.4 with a $\rho$ of 0.3. This result is in accordance with prior literature, which has estimated values between 1 and 3 for this parameter, as reported by Katz and Autor (1999) in their literature review.

Panel B of table 7 reports estimates of the labor supply. Overall, these estimates suggest that workers, in general, prefer cities with higher wages, lower rents, higher population, and lower low-skill service prices. A 1% increase in local population increases the local high- and low-skill working population by 4%. At the same time, increasing high- and low-skill wages by 1% increases the high- and low-skill working population by 3.6%. However, increasing low-skill non-tradable and housing prices by 1% decreases the high- and low-skill working population by 3.5%. Given that I assume that the goods are produced jointly according to a Cobb-Douglass function, I infer the expenditure shares from workers' revealed preference in the trade-off between wages, rent and service prices across cities. The indirect utility value of rent and non-tradable service prices measured in terms of wage units represents the share of expenditure for the tradable good. I estimate this to be approximately 50%. This value is lower than that found by Diamond (2016) and Albouy (2008). However, it must be noted that the time periods and identification strategies used here are not the same as those used in these prior papers. In the Appendix, I show how the model predicts wage convergence and its decline using a Bartik shock rather than an SBTC shock in section A.3.3.

### 5.2.2 Estimation Robustness

To assess whether these parameter estimates are sensitive to different ways of measuring the model's variables or instruments, I conduct various robustness checks. The two principal robustness checks are 1) Running the estimates with a different sample; 2) Running the estimates including only the MSAs included in the pre-1980 Census data. As a further robustness check, I obtain estimates for 1940-1980 looking at states instead of MSAs. While there are fewer states than MSAs, which leads to less variation, the states are consistent across time. Therefore, these estimates are more representative than the national average. The results of the state-level analysis, however, are quite consistent with those of the MSA-level analysis.

### 5.2.3 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 8. In particular, I borrow the share of non-tradable services with respect to housing, $\nu$ from Serrato and Zidar (2014). To include housing in the model with non-homothetic preferences, I also include a subsistence level of housing, $\bar{O}$, from Ganong and Shoag (2015), which is set to match the Engel curve for housing. To complete the housing sector, a value for the elasticity of housing, $\mu$, is needed. This elasticity is also borrowed from Ganong and Shoag (2015). It is chosen to generate a 1-to-1 relationship between log prices and log per capita incomes in order to match the relationship from the data, and it 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 $\gamma_2$, which relates to the relationship between growth and population distribution, from Desmet et al. (2015).[28] The parameters of the migration cost function, which is exponential, are different for high-skill and the low-skill workers. I borrow these estimates from Notowidigdo (2011), which uses an identification strategy based on Bartik instruments. 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 2 types of workers.[29]

## 6 Model Simulation and Counterfactuals

In this section, I first describe how the estimation is conducted. Second, I provide more details as to how the numerical computation of the equilibrium is achieved. Third, I show how the model matches non-targeted moments in the decline in $\beta$-convergence. Specifically, the model fits well the decline in spatial convergence for high-skill workers. Fourth, I conduct a quantitative decomposition of each

[28] I am currently working on estimating $\gamma_2$ from the productivity terms $\xi_{kjt}$ that I obtain from the GMM estimation.

[29] Extending the migration cost algorithm is not the primary focus of this paper and, therefore, it is left for future work.

mechanism's effect on the decline in convergence. Fifth, I investigate whether the model matches other non-targeted moments such as the "Great Divergence" of skills, the decline in migration and the increase in wage dispersion.

The estimation procedure above allows us to obtain values for all 10 model parameters, the initial productivity terms and SBTC. Now, we can compute the equilibrium of the model by solving a system of equations for every period $t$, incorporating the productivity values from the previous period.

The model can be reduced to 7 equations, as shown in the Appendix. Given that the analysis includes 240 cities, the iteration procedure contains 1680 equations for each time period $t$. The equilibrium conditions correspond to equations 17, 18, 20, 6, 10 and 11. Because of the large number of cities, the problem is highly dimensional. An extra complication of the model is the endogenous agglomeration effects, which could induce the system of equations to explode. However, the estimates respect the restrictions imposed by the low-skill service sector and the system is stable. As a robustness test, I conduct a sensitivity analysis and check whether varying the parameters changes the results substantially and whether the system maintains wage convergence. More details about these conditions can be found in Appendix section B.1.

## 6.1 Model vs. Data

I now show that the model provides a good fit to the patterns documented in the previous subsections. Specifically, with the wages for high and low-skill workers produced by the model, I run the same regression within my model as I did with the data in section 3. In fact, also in the model I estimate the $\beta_k$-convergence using the regression proposed in section 3 following Baumol (1986)

$$\frac{w_{kjt} - w_{kj\tau}}{T - \tau} = \alpha + \beta_k w_{kj\tau} + \epsilon \tag{33}$$

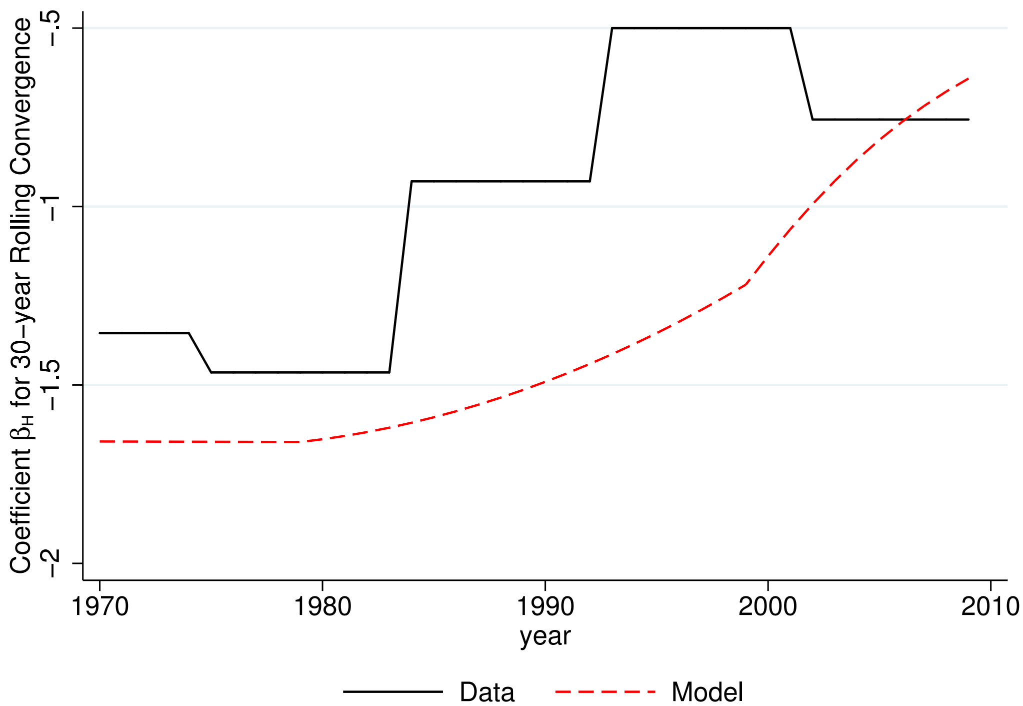
where $\Delta w_{kjt}$ is the annual wage growth between time $t$ and $\tau$, $k$ is the skill-type, which may be $k \in \{H, L\}$ in location $j$ and $w_{kj\tau}$ is the wage in the initial period $\tau$ for type $k$ in location $j$. In this context, the difference between $t$ and $\tau$ is equal to 20 years.

Then, I show that the model provides a good fit to the convergence patterns documented above. Specifically, I construct the evolution of the $\beta$-convergence for average wages and for wages of high- and low-skill workers, following equation 33.

In Figure 6, I plot the estimated $\beta$-convergence from the model and from the data to compare them. I average out the estimates of $\beta_H$- and $\beta_L$-convergence weighted by the population shares. The estimates from the data look like a step function since the data are available only every 10 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 .005% points.

Figure 7 compares the $\beta_H$-convergence rates over time both in the data and in the model. The estimates are very close over time. Also, the first row of table 9 reports the correlation between the data and the model. The model performs very well in fitting wage convergence pattern in the non-targeted moment and the decline of convergence for the high-skill group. Specifically, the correlation is 0.92. The match is not as good for the wage convergence rate among low-skill workers, which is reported in figure 8, because of the degree of substitutability between H and L. I run simulations in which I increase the substitutability between the two labor inputs and the $\beta_L$-convergence fits the data more closely.

Figure 6: Model Matching the Data on Wage Convergence

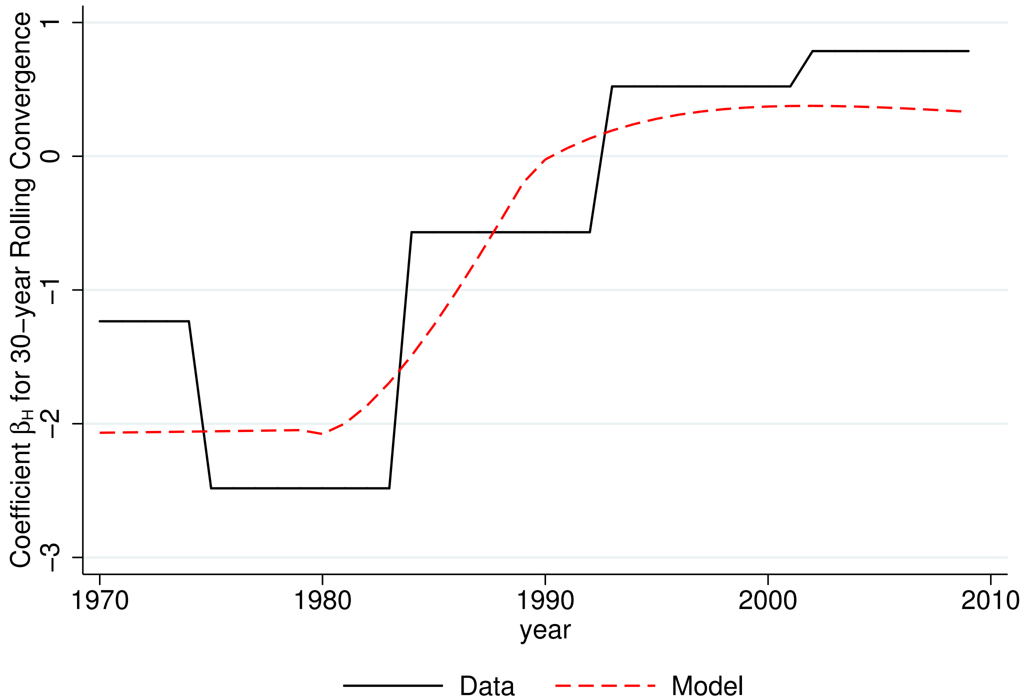



Note: This figure shows a rolling estimate of the $\beta$-convergence over 30 years. The solid line is the data for which we have observations every 10 years (that is why it looks like a step function), while the dashed line is the estimate of the $\beta$-convergence from the model, for which we can compute a yearly estimate.

## 6.2 Quantitative Decomposition

Once checked that the model fits the data, I calculate several counterfactual scenarios for the $\beta$, $\beta_H$ and $\beta_L$ convergence rate, which allow us to assess the quantitative contributions of each of the model's mechanisms. I use the calibrated model as a baseline specification from which to depart and construct counterfactuals. Specifically, I proceed stepwise, sequentially "turning off" each component of the model that contributes to the decline in wage convergence over time. This

Figure 7: Model Matching the Data on High-Skill Wage Convergence

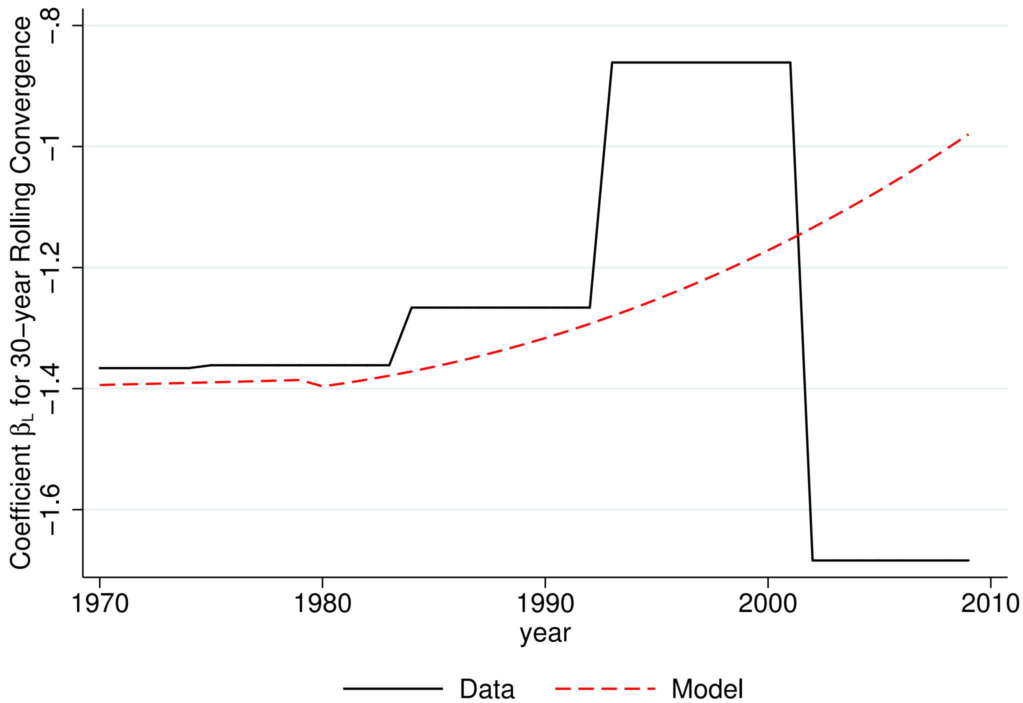



Note: This figure shows a rolling estimate of the $\beta_H$-convergence over 30 years. The solid line is the data for which we have observations every 10 years (that is why it looks like a step function), while the dashed line is the estimate of the $\beta$-convergence from the model, for which we can compute a yearly estimate.

decomposition for $\beta_H$-convergence, which is the one driving the end of convergence, is illustrated in figure 11.

My counterfactual of interest is comparing estimates of $\beta_H$ in 2010 in the baseline model with the estimates that I obtain once I "turn off" the mechanisms. Starting from the baseline convergence rate that is equal to 0.33% a year, in plot (a), I remove migration costs by setting $m_{H2}$ and $m_{L2}$ to 0. The convergence rate decreases by .09 percentage points to .24% per year. In plot (b), I also remove the housing sector, setting $\bar{O}$. As a result, the $\beta$-convergence rate for high-skill workers on the right declines by .04 percentage points. In plot (c), I also remove the agglomeration forces. Specifically, I set $\gamma^H$, $\gamma^L$, $\gamma$ to 0. The $\beta$-convergence rate decreases substantially, by .776 percentage points compared to plot (b). Finally, in plot (d), I remove SBTC ($S_H$ and $S_L$ in the model) by setting $\lambda^H$ and $\lambda^L$ to be equal to 0. The $\beta$-convergence between 1980 and 2010 is estimated to be -1.5% a year. Interestingly, this result suggests that without SBTC, the wage convergence rate would be the same as that observed between 1940 and 1980. Overall, the main finding is that the

Figure 8: Model Matching the Data on Low-Skill Wage Convergence



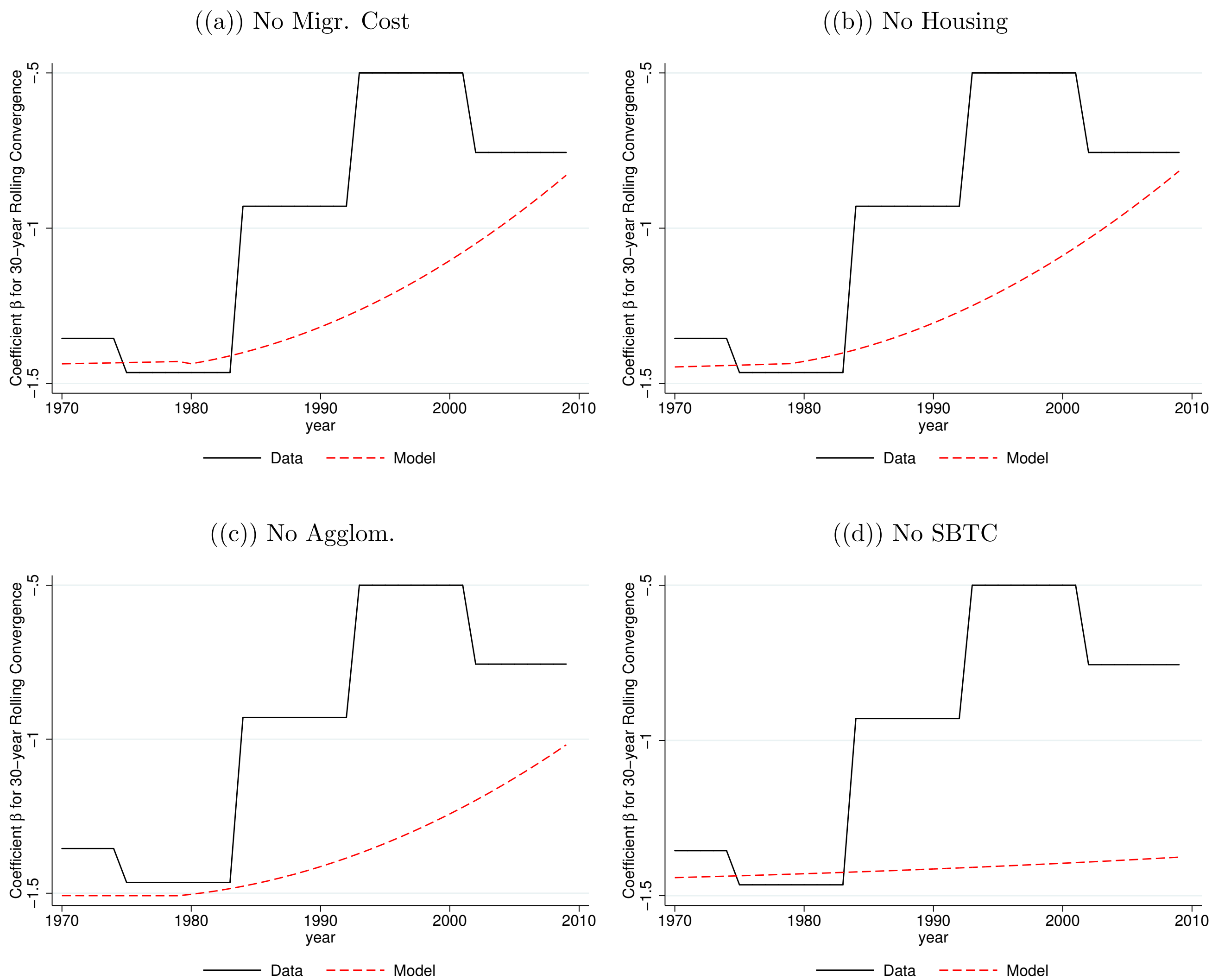

Note: This figure shows a rolling estimate of the $\beta_H$-convergence over 30 years. The solid line is the data for which we have observations every 10 years (that is why it looks like a step function), while the dashed line is the estimate of the $\beta$-convergence from the model, for which we can compute a yearly estimate.

bulk of the decline in convergence after 1980 can be attributed to SBTC.

The results of this counterfactual analysis demonstrate that the convergence rate with endogenous productivity channels and SBTC would be about 1.2% a year. Instead, if I shut down the productivity channel, nominal wage convergence would be about 1.1% a year. The model fit predicts that the wage convergence rate for the low-skill did slightly reduce over time. This is because there are complementary forces that cause low-skill wages to increase. I run simulations in which $\rho$ is larger, indicating higher substitutability between high-skill and low-skill. The results suggest that the higher $\rho$, the lower the correlation between $\beta_H$ and $\beta_L$ convergence.[30]

[30]In another counterfactual scenario, I generate real wage convergence. The model has the advantage of allowing for service prices to be different across space. Therefore, we are able to calculate real convergence. Unfortunately, there are no data with which to compare this result because price indicators by MSA are not reported in 1980 or before.

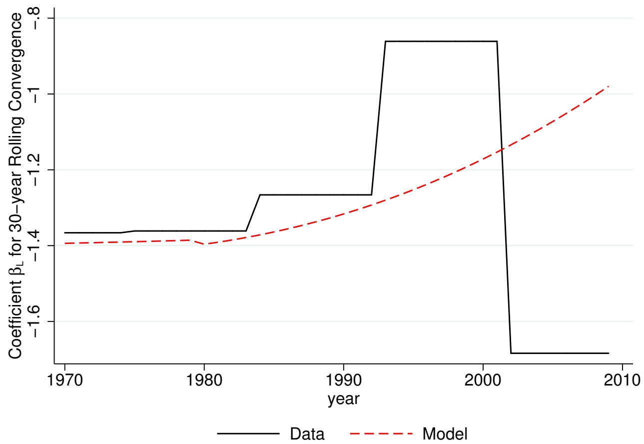
Figure 9: Quantitative Decomposition of Wage Convergence



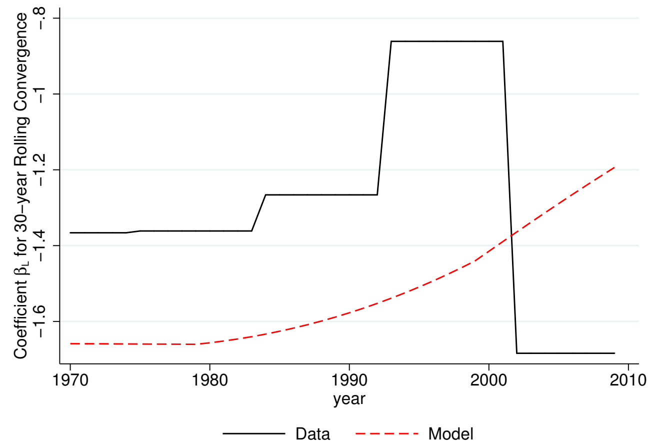

Note: This figure shows counterfactual exercises in which I shut down cumulatively in order migration cost in subplot (a), housing in subplot (b), agglomeration forces in subplot (c) and finally SBTC in subplot (d).

Figure 10: Quantitative Decomposition of Low-Skill Wage Convergence

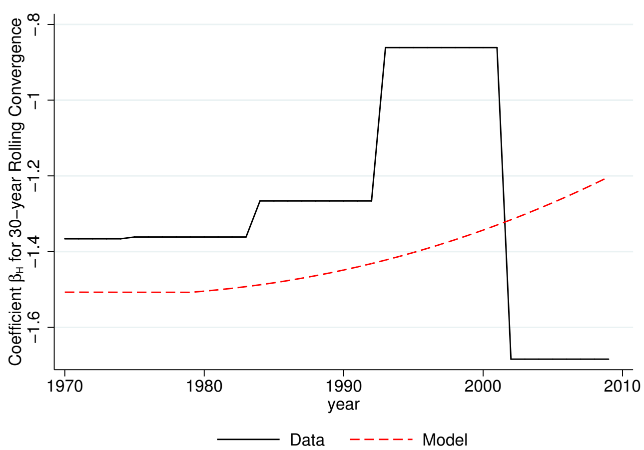
((a)) No Migr. Cost



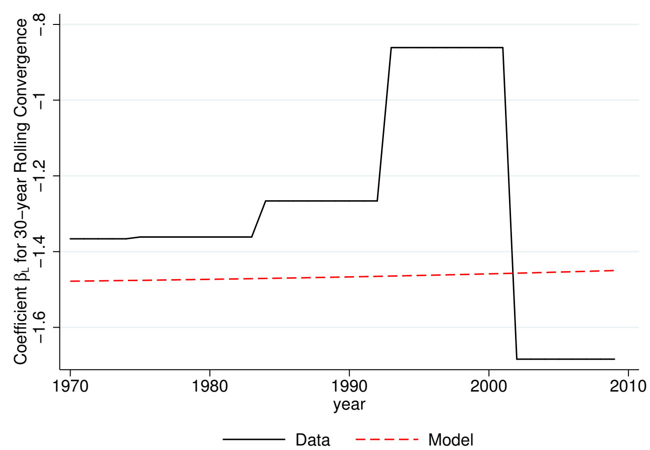

((b)) No Housing




((c)) No Agglom.



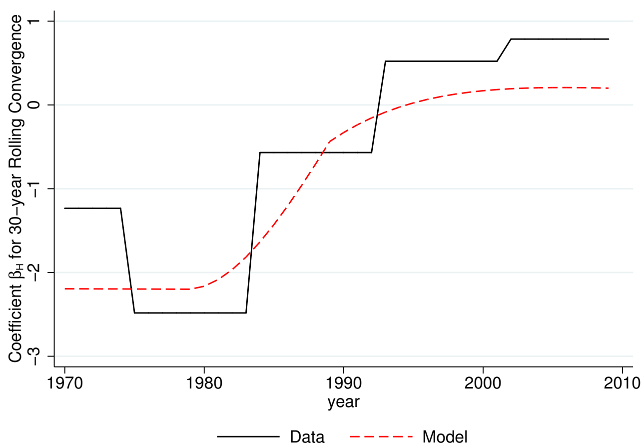

((d)) No SBTC



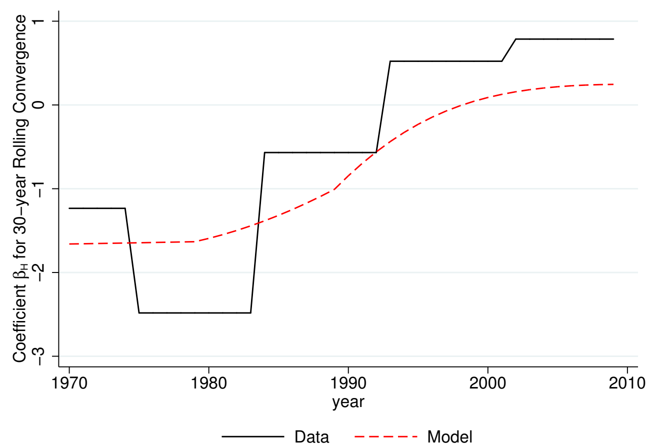

Note: This figure shows counterfactual exercises in which I shut down cumulatively in order, migration cost in subplot (a), housing in subplot (b), agglomeration forces in subplot (c) and finally SBTC in subplot (d).

Figure 11: Quantitative Decomposition of High-Skill Wage Convergence

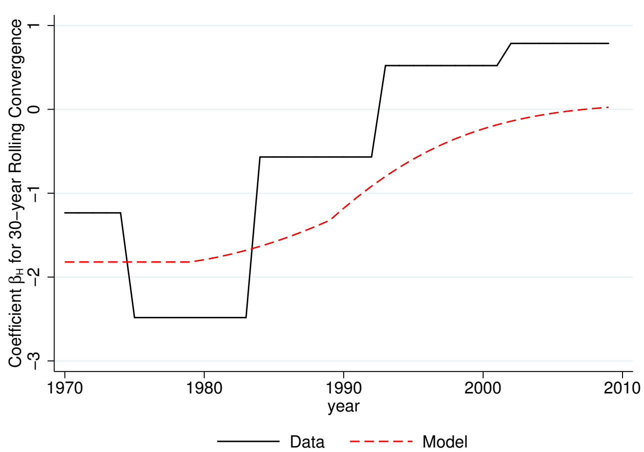
((a)) No Migr. Cost



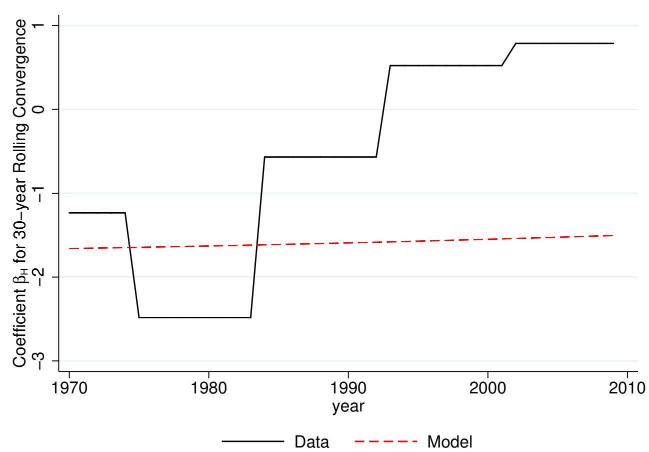

((b)) No Housing




((c)) No Agglom.




((d)) No SBTC




Note: This figure shows counterfactual exercises in which I shut down cumulatively in order, migration cost in subplot (a), housing in subplot (b), agglomeration forces in subplot (c) and finally SBTC in subplot (d).

## 6.3 Wage Dispersion Increase Over Time

Hsieh and Moretti (2015) show that wage dispersion across U.S. cities increased substantially between 1964 and 2009. Does this model match this increase in wage dispersion over time? The model suggests that wage dispersion in the U.S. has increased substantially over the last 30 years. This accords with the empirical findings in Hsieh and Moretti (2015), which finds that wage dispersion almost doubles between 1964 and 2009. My model supplements this finding by predicting differences in wage dispersion between high-skill and low-skill workers. In particular, the model predicts that wage dispersion increased solely in the high-skill group. The model predicts a wage dispersion increase of 50% while the data suggest a wage dispersion increase of almost 100%. Although the model does not replicate the full increase in wage dispersion in the US, it predicts part of it. Hsieh and Moretti (2015), using a framework where the dispersion forces are housing and amenities, suggest that the main reason for the increase in wage dispersion was the housing sector. To control whether in my framework, where besides housing and amenities, there are other dispersion forces such as SBTC, I conduct a counterfactual analysis to check what are the determinants of wage dispersion over time. If we look at table 10, we can see that the model produces a wage dispersion increase for high-skill workers but not a significant one for the low-skill group. Taking the average of the two groups, the increase in wage dispersion matches the findings from Hsieh and Moretti (2015). From the second row of the table onwards 10, I report the results of a quantitative decomposition run in the same fashion as the one reported in section 6.2. The results suggest that even in this case, the biggest chunk of the decline in wage dispersion is attributable to SBTC, rather than housing, as found by Hsieh and Moretti (2015) in a context where SBTC was not present.

## 6.4 The Convergence and Divergence of Skill ratio over time

What happened to the distribution of high and low-skill workers over time and across space? Moretti (2003), Berry and Glaeser (2005), Diamond (2016) and Moretti (2012) show that the skill ratio of workers between 1980 and 2010 was diverging across cities. Specifically, Diamond (2016) finds that a 1% greater skill ratio in 1980 is associated with a .17% larger increase in the MSA's college employment ratio from 1980 to 2000. Moretti (2012) coins the term "Great Divergence" to emphasize how the skill ration has diverged over space. But what happened to skill ratio over time? Was it diverging or converging before 1980, when U.S. cities were experiencing wage convergence? To answer this question, I look at the convergence rates of the skill ratio over time, not just between 1970 and 2000 as in the existing literature (Berry and Glaeser 2005, Moretti 2012, Diamond 2016). I estimate the following specifications exactly as I did with the data:

$$\log\left[\frac{H_{jt}}{L_{jt}} - \frac{H_{j\tau}}{L_{j\tau}}\right] = \alpha + \beta \cdot \log\frac{H_{j\tau}}{L_{j\tau}} + \epsilon \tag{34}$$

Table 5 shows the results of the specification above. In Panel A, I report the results with a difference between $t$ and $\tau$ of 10 years. While in Panel B, the difference is set to be 20 years. As we can see in Panel A, the estimated coefficients are negative and statistically significant until 1970. They are not significant for 1970-1980 and 1980-1990, then, 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 .07% and .04%, respectively between 1990-2000 and 2000-2010. In Panel B, the results are quite similar, however, in column (1), the coefficient is positive and statistically significant. That coefficient is actually calculated between 1940 and 1970 since data for 1960 are not available. Therefore, it could be the case that in this 30 year time span, the results are reversed for other reasons. But, the coefficient between 1950 and 1980 is negative and statistically significant as expected. Specifically, a 1% increase in the college ratio in 1950 decreases the change in college ratio between 1950 and 1980 by .32%. If we compare the results in table 5 with the estimate derived from the model in table 11, we notice that the estimates go in the same direction and the magnitudes are very similar. Specifically, the coefficient between 1940 and 1950 is -.218 for the data and -.245 for the model. The coefficient between 2000 and 2010 is 0.04 for the data and 0.0826 for the model and both are statistically significant. The model generates estimates that are more precise than the data. Overall, my model is able to reproduce another non-targeted moment, which suggests that it is able to reproduce 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.

### 6.5 Decline in Gross Migration Flows over time

The bulk of this paper aims at understanding the causes of the decline in the regional convergence in wages. However, the decline in interstate migration is another important structural change that happened in the U.S. 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 (2015) provide and test a theory of reduction in the geographic specificity of occupations coupled with information technology and inexpensive travel. They suggest 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. If, moreover, the technological shock persists over time, then this effect will become even stronger, decreasing migration even further. For instance, suppose that a high-skill worker lives in San Francisco in the nineteen eighties. When the technology shock arrives, the high-skill worker will have less incentive to move out because San Francisco becomes even more attractive to him. Another high-skill 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 low-skill workers. Figure 14 shows that the model matches the data for the migration rate of high-skill workers reasonably well, but does not a reproduce sufficient decline in migration among low-skill workers. If we look at the correlation between the data and the model, shown in table 12, we see a higher correlation for high-skill migration.

#### 6.5.1 Sorting of Rich Cities and High-Skill Migrants

In section 3, I document that high-skill workers are moving more and more to high-skill MSAs. In this section, I check how the model matches this feature of the data. The model, however, does not distinguish migrants from non-migrants. But, it allows calculating migration rates by taking the differences of the population in a MSA over time. To check how the model matches the data, I generate data on changes in the population of high-skill $H$ in MSA $j$ and on average wages of MSA $j$, then, I run the following regression:

$$\Delta H_{jt} = \alpha + \sum_{t=1941}^{2010} \delta_t^H \ln_{Hjt} + \epsilon_t \tag{35}$$

In figure 15, I plot the estimates of $\delta_t^H$for each year. The results show that the MSA with higher average wages, had a decrease in the number of high-skill migrants between 1940 and 1980. Specifically, a 1% increase in wages will generate approximately 1.4% decrease in the high-skill 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 will increase the number of high-skill migrants by 2%.

## 7 Other Potential Explanations

There are several potential explanations that are complementary to SBTC and agglomeration story. In this section, I focus on housing regulations, industry composition and firms' location decisions and *Right to Work Laws*.

### 7.1 Housing Regulation

Ganong and Shoag (2015) provide a story of housing prices that suggests that U.S. states where the housing prices increased the most are also the ones where the migration declined. Hence, because migration increases convergence, the decline in migration to this areas, that are found to be also the richest, declined the income convergence rate. As stated in their paper, the housing prices and SBTC story can be complementary. For this reason, in order to decide how to disentangle between the two of them, I introduce a non-tradable housing sector in the model to compare the housing effects with my key mechanisms.

Moreover, as an extra test, I conduct a simple empirical test. I show that even in the areas where the housing restrictions are high, there is a strong difference in the convergence rate of wages for the high-skill group and the low-skill group. I construct figure 2 only for MSA that were in states where the housing prices went up dramatically because of the high housing regulations. If I look at figure 16, I can see that the effect of this on the decline in income convergence looks quite similar to the one without any restriction. From this, I can conclude that there is room also for a story of skilled-biased technical change also in the group of states where housing prices were high.

## 7.2 Innovation and Financial Sector

Another potential and complementary explanation is that the 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 sectoral innovation rather than skill composition, I estimate conditional convergence in wages between 1980 and 2010 controlling for the IT sector.[31] The results reported in table 13 suggest 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 on wages in 1980 becomes positive and statistically significant. In column C, I add a control for college degree, and the coefficient on initial wages in 1980 increases in magnitude. This evidence suggests that including sectoral differences in technological intensity has the effect of amplifying the decline in spatial convergence. The framework developed above takes into account these sectoral differences by including a high-skill and low-skill sector.

In addition to sectoral innovation shifts, changes in firms' relocation decisions over time may have contributed to the decline in wage convergence. More skilled firms may have begun to move to richer places, reversing their earlier tendency to move to poorer cities to take advantage of lower costs. In order to investigate whether firms' location decisions have changed over time requires firm-level data. Faberman and Freedman (2016), using longitudinal establishment data for the U.S. during the years 1992-97, do not find that spillover is important for firms' decision to locate in urban areas rather than other areas. Unfortunately, data on firms location back to 1940 are not available. In this regard, I use publicly available data at the industry level to test whether more-skilled occupations have become increasingly concentrated in more-skilled cities over time. If this is the case, it may imply that in addition to sorting of high-skill workers into high-skill cities, there is also sorting of high-skill firms into high-skill cities. To test this hypothesis empirically, I

[31] I define IT sector by looking at the codes of the IND1990 variable in the IPUMS dataset and selecting industries that are more technology-oriented.

run the following regression to obtain the marginal effects by decade

$$\text{Skill concentration}_{kjt} = \alpha + \sum_{t=1950}^{T} \beta_t \left(\frac{H_{jt}}{L_{jt}}\right) \cdot + \phi_t + \phi_j + \epsilon_{kjt} \tag{36}$$

where $k$ is the industry, $j$ is the MSA and $t$ is time, $fe_{time}$ are time fixed effects and $fe_{MSA}$ 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 there are in industry $k$ in location $j$ at time $t$. This hypothesis is confirmed in the data. In figure 17 I plot the coefficient $\beta_t$ over time. We can see that over time, being in a more skill-concentrated MSA becomes more strongly correlated with skill concentration at the industry level. This suggests evidence of sorting not just of workers but also sorting of industries and thus, firms.

### 7.3 Right to Work Laws

26 states, mostly in the Southern and Western U.S. have passed *Right to Work Laws* since 1940. These laws permit workers to work without having joining a union. The *Right to Work Laws* may have had a spatial effect of increasing the wages of low-skill workers in the states where they were implemented. In fact, Holmes (1998) documents that state policies play a role in the location of industry. However, only 26 states have adopted right to work laws and, moreover, if we look at the timeline of the laws' passage in figure 18, we can see that the majority of the states passed these laws in the 1950s and 1960s, long before the secular decline in wage convergence.

## 8 Conclusions and Potential Extensions

In this paper I document that the decline in wage convergence among MSAs observed after 1980 was largely due to the decline in wage convergence among high-skill workers, whereas wage convergence among low-skill workers did 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 focusing on the role of the interaction of SBTC and agglomeration effects.

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 declined only among high-skill workers after 1980; 1) The skill premium is higher in educated cities after 1990; 2) Over time, high-skill workers begin to migrate relatively more than low-skill workers to educated cities. Decomposing this relationship, I find that this pattern is stronger for high-skill workers than for low-skill workers.

I estimate the model using a GMM estimation procedure, employing an SBTC shock and housing regulations as sources of exogenous variation. The model estimates are consistent with the conclusions of prior literature. Using model estimates to calibrate some parameters and calibrating other parameters from the existing literature, I compute the equilibrium of the model. The calibrated model provides a good fit to the data and shows that both SBTC and agglomeration effects play an important role in explaining the decline in wage convergence among the high-skill workers. The main findings suggest that SBTC explains approximately 80% of the decline in cross-MSA wage convergence in the U.S. after 1980.

Moreover, the model matches non-targeted moments including the increase in wage dispersion over the last 40 years documented by Hsieh and Moretti (2015) and the "Great Divergence" in skills addressed by Moretti (2012).

This paper is one of the first to study the interaction of agglomeration effects and SBTC. Moreover, to be best extent 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 U.S. regions and increased income inequality for different levels of skills may have important policy implications especially for the regions which are not able to grow as the richer regions. Dealing with sustaining the growth in the richest MSAs and arresting decline in poorer MSAs is an important challenge to face for policy makers. Moreover, understanding regional inequality contributes to understand the skill premium between high-skill and low-skill workers. Despite the fact that the mechanism proposed cannot explain entirely the stop of convergence and the change in the skill premium, because of the complexity of the phenomenons, I also think that it capture an important component of them and it may also have external validity when I ask why cross-country convergence does not hold as Barro and Sala-i-Martin (1992) show.

The framework of this paper is flexible enough that can be extended to perform several types of analysis, including a cross-country analysis. I have done some preliminary work that suggests that there was a decline in regional income convergence across 155 countries after 1980. In future work, I will use my analytical framework to investigate this decline.

## References

Daron Acemoglu and David Autor. Skills, Tasks, and Technologies: Implications for Employment and Earnings. 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 2008.

Treb Allen and Costas Arkolakis. Trade and the Topography of the Spatial Economy. (19181), June 2013.

David Autor and David Dorn. The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market. 2013.

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.

R.J. Barro and X. Sala-I-Martin. Convergence. 1991.

R.J. Barro and X. Sala-I-Martin. Technological diffusion, convergence, and growth. (5151), June 1995. 10.3386/w5151.

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? 2014.

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. (13111), November 2013.

Martin Beraja, Erik Hurst, and Juan Ospina. The aggregate implications of regional business cycles. (21956), February 2016. 10.3386/w21956.

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.

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

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? (19010), May 2013. 10.3386/w19010.

Donald R. Davis and Jonathan I. Dingel. The Comparative Advantage of Cities. 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.

Giles Duranton and Diego Puga. Micro-foundations of urban agglomeration economies. (9931), September 2003. 10.3386/w9931.

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

Peter Ganong and Daniel Shoag. Why Has Regional Convergence in the U.S. Stopped? (rwp12-028), June 2015.

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.

Claudia Goldin and Robert A. Margo. The great compression: The wage structure in the united states at mid- century. *The Quarterly Journal of Economics*, 107(1):1–34, 1992.

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 David H. Autor. *Changes in the Wage Structure and Earnings Inequality*, pages 1463–1555. 1999.

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.

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. (9641), April 2003. 10.3386/w9641.

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

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

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. 125(3):1253–1296, 2010. 10.1162/qjec.2010.125.3.1253.

Sam Schulhofer-Wohl and Greg Kaplan. Understanding the Long-Run Decline in Interstate Migration: Online Appendix. (725), June 2015.

Juan Carlos Suárez Serrato and Owen Zidar. Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. (20289), July 2014. 10.3386/w20289.

# A Appendix

## A.1 Tables

Table 1: Wage Convergence Rates

Panel A: Convergence Rate

| | Population Weighted (pw) | | Unweighted | |
|---|---|---|---|---|
| | $\Delta w^{40-80}\_pw$ | $\Delta w\_pw^{80-10}$ | $\Delta w^{40-80}$ | $\Delta w^{80-10}$ |
| Log wages, 1940 | -0.0112*** | | -0.0144*** | |
| | (-10.90) | | (-16.81) | |
| Log wages, 1980 | | -0.0000389 | | -0.00852* |
| | | (-0.02) | | (-2.57) |
| Constant | 0.000360* | 0.00145*** | -1.37e-09 | -0.0000229 |
| | (2.29) | (4.90) | (-0.00) | (-0.09) |

$t$ statistics in parentheses

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

Panel B: Convergence Rate by College Degree

| | 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 $\beta$-convergence plotted in figure 1 and 2. In Panel A, I report the estimate of the $\beta$ coefficient for the whole sample underlying figure 1. In column (1), there are $\beta$ estimates for 1940-1980, and the observations are population weighted. Column (2) reports same estimation but for 1980-2010. In columns (3) and (4), the estimations are not population weighted. In Panel B.1, I report the estimates of the $\beta$-convergence corresponding to figure 2. In column (1), I report the estimate for non-college graduates for 1940 and 1980; in column (2), for college graduate in the same time period. In columns (3) and (4), the estimates are once again for the two groups, but for the 1980-2010 time span. 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. ***, **, and *, represent statistical significance at 0.001, 0.01 and 0.05, 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: The table reports the coefficients of OLS regressions. The dependent variable is the skill premium measured as the difference between the log wages of college graduates and non-college graduates. The only difference between column (1) and column (2) is that I control for population in level in column (1). The t-statistics are presented in parentheses. Observations are clustered at the state level. ***, **, and *, represent statistical significance at 0.001, 0.01 and 0.05, respectively.

Table 3: Skill Premium by College Ratio of Cities by Year

| | (1) Skill Premium | | (2) Skill Premium | |
|---|---|---|---|---|
| Coll. Ratio in 1964 | -0.0813 | (-0.74) | -0.102 | (-0.94) |
| Coll. Ratio in 1965 | -0.0841* | (-2.02) | -0.0942* | (-2.24) |
| Coll. Ratio in 1966 | -0.102** | (-2.63) | -0.112** | (-2.87) |
| Coll. Ratio in 1967 | -0.102** | (-2.71) | -0.112** | (-2.93) |
| Coll. Ratio in 1968 | -0.108** | (-2.69) | -0.119** | (-2.93) |
| Coll. Ratio in 1969 | -0.1000* | (-2.39) | -0.111** | (-2.63) |
| Coll. Ratio in 1970 | -0.0943* | (-2.18) | -0.105* | (-2.41) |
| Coll. Ratio in 1971 | -0.111** | (-2.61) | -0.122** | (-2.83) |
| Coll. Ratio in 1972 | -0.108* | (-2.50) | -0.119** | (-2.73) |
| Coll. Ratio in 1973 | -0.124** | (-2.88) | -0.135** | (-3.12) |
| Coll. Ratio in 1974 | -0.124** | (-2.76) | -0.136** | (-2.99) |
| Coll. Ratio in 1975 | 0.00840 | (0.24) | -0.00302 | (-0.08) |
| Coll. Ratio in 1976 | -0.00465 | (-0.14) | -0.0158 | (-0.46) |
| Coll. Ratio in 1977 | 0.00663 | (0.20) | -0.00360 | (-0.10) |
| Coll. Ratio in 1978 | 0.00114 | (0.03) | -0.00914 | (-0.26) |
| Coll. Ratio in 1979 | 0.0104 | (0.30) | 0.00000428 | (0.00) |
| Coll. Ratio in 1980 | 0.00747 | (0.22) | -0.00328 | (-0.09) |
| Coll. Ratio in 1981 | -0.00757 | (-0.22) | -0.0181 | (-0.49) |
| Coll. Ratio in 1982 | -0.00196 | (-0.06) | -0.0127 | (-0.35) |
| Coll. Ratio in 1983 | -0.0179 | (-0.51) | -0.0290 | (-0.78) |
| Coll. Ratio in 1984 | -0.0282 | (-0.81) | -0.0395 | (-1.05) |
| Coll. Ratio in 1985 | -0.0417 | (-1.82) | -0.0404 | (-1.72) |
| Coll. Ratio in 1986 | -0.0404 | (-1.84) | -0.0392 | (-1.78) |
| Coll. Ratio in 1987 | -0.0459* | (-2.05) | -0.0447* | (-1.98) |
| Coll. Ratio in 1988 | -0.0370 | (-1.65) | -0.0358 | (-1.61) |
| Coll. Ratio in 1989 | -0.0404 | (-1.76) | -0.0398 | (-1.76) |
| Coll. Ratio in 1990 | -0.0368 | (-1.58) | -0.0353 | (-1.54) |
| Coll. Ratio in 1991 | -0.0453* | (-2.03) | -0.0439* | (-1.97) |
| Coll. Ratio in 1992 | -0.0844* | (-2.31) | -0.0818* | (-2.26) |
| Coll. Ratio in 1993 | -0.0993** | (-2.66) | -0.0966** | (-2.61) |
| Coll. Ratio in 1994 | -0.0894* | (-2.36) | -0.0868* | (-2.32) |
| Coll. Ratio in 1995 | 0.0422 | (1.39) | 0.0658* | (2.20) |
| Coll. Ratio in 1996 | 0.0293 | (0.96) | 0.0527 | (1.75) |
| Coll. Ratio in 1997 | 0.0309 | (1.03) | 0.0543 | (1.83) |
| Coll. Ratio in 1998 | 0.0197 | (0.63) | 0.0439 | (1.42) |
| Coll. Ratio in 1999 | 0.0121 | (0.37) | 0.0367 | (1.14) |
| Coll. Ratio in 2000 | 0.0122 | (0.40) | 0.0366 | (1.26) |
| Coll. Ratio in 2001 | -0.000632 | (-0.02) | 0.0222 | (0.69) |
| Coll. Ratio in 2002 | 0.0116 | (0.36) | 0.0352 | (1.14) |
| Coll. Ratio in 2003 | 0.0121 | (0.36) | 0.0356 | (1.13) |
| Coll. Ratio in 2004 | 0.0200 | (0.62) | 0.0434 | (1.35) |
| Coll. Ratio in 2005 | 0.122** | (2.65) | 0.147*** | (3.31) |
| Coll. Ratio in 2006 | 0.109* | (2.41) | 0.134** | (3.05) |
| Coll. Ratio in 2007 | 0.118* | (2.46) | 0.145** | (3.13) |
| Coll. Ratio in 2008 | 0.116* | (2.23) | 0.143** | (2.88) |
| Population | | | 0.0116*** | (6.91) |
| Time fixed effects | yes | | yes | |
| N | 1555 | | 1555 | |

Note: The table reports the coefficients of OLS regressions. The dependent variable is the skill premium measured as the difference between the wages of college graduates and non-college graduates. T-statistics are presented in parentheses, and observations are clustered at the state-level. ***, **, and *, represent statistical significance at 0.001, 0.01 and 0.05, respectively. Column (2) is identical to column (1) with the exception of controlling for population. These estimates are run with CPS data.

Table 4: 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 | |

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

Table 5: $\Delta \frac{H}{L}$ 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 reports 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 of $\Delta \frac{H}{L}$. Standard errors are in brackets. ***, **, and *, represent statistical significance at 0.001, 0.01 and 0.05,respectively.

Table 6: First Stage Estimates of Models for Routine Occupation Share Measures

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

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

Table 7: Model Estimates for 1940-2010

| Panel A: Labor Demand (1940-2010) | | |
|---|---|---|
| Moments | Parameter | Estimates |
| $E[\Delta\xi_{kjt}\Delta Z_{jt}]=0$ | endogenous productivity $H$ workers: $\gamma^H$ | 0.285*** |
| | endogenous productivity $L$ workers: $\gamma^L$ | 0.004 |
| | endogenous population productivity: $\gamma$ | 0.312*** |
| | $\frac{1}{1-\rho}$ elasticity of substitution between $H$ and $L$: $\rho$ | 0.30*** |
| Panel B: Labor Supply (1940-2010) | | |
| $E[\Delta A_{kjt}\Delta Z_{kjt}]=0$ | Elasticity to wages: $\beta^w$ | 3.6*** |
| | Elasticity to local prices: $\beta^s$ | -3.5*** |
| | Elasticity to population: $\gamma$ | 3.1*** |

Note: In this table, I report the moments and the estimates of the model. In Panel A, I report estimates of labor demand. In Panel B, I report estimates of the labor supply.***, **, and *, represent statistical significance at 0.001, 0.005, and 0.01 levels, respectively.

Table 8: Externally calibrated Parameters

| Parameter | Value | Literature |
|---|---|---|
| Share of Non-Tradable $NT$ : $\nu$ | 0.2 | Serrato and Zidar (2016) |
| Subsistance level of Housing: $\bar{O}$ | 0.25 | Ganong and Shoag (2015) |
| Elasticity of Supply Housing: $\mu$ | 0.4 | Ganong and Shoag (2015) |
| Share of technology: $\gamma^2$ | 0.99 | Desmet et al. (2016) |
| Migration costs: $\sigma^L$ and $\beta^L$ | -.065 and -.861 | Notowidigdo (2013) |
| Migration costs: $\sigma^H$ and $\beta^H$ | -.066 and -1.044 | Notowidigdo (2013) |

Table 9: $\beta_H$-Convergence: Correlation Data and Model

| | (1) | |
|---|---|---|
| | Corr. Data and Model | |
| | mean | count |
| Corr. Data and Model | .9200265 | 40 |
| No Migr. Costs | .9153843 | 40 |
| No Housing | .910886 | 40 |
| No Agglomeration | .9015143 | 40 |
| noSBTC2betaH | -1.589833 | 40 |
| Observations | 40 | |

Note: This table reports the correlation between the convergence rate produced by the model and the convergence rate produced by the data for high-skill workers over time.

Table 10: Wage Dispersion Increase in the Last 30 Years

| | (1) Wage DispersionH mean |
|---|---|
| Wage Dispersion | 2.615298 |
| No Migr. Costs | 2.22446 |
| No Housing | 2.22446 |
| No Agglomeration | 2.104676 |
| No SBTC | .3861664 |
| Observations | 1 |

| | (1) Wage DispersionL mean |
|---|---|
| Wage Dispersion | 1.106282 |
| No Migr. Costs | .4199571 |
| No Housing | .4199571 |
| No Agglomeration | .7551879 |
| No SBTC | .3419744 |
| Observations | 1 |

Note: The first row of the table on the left reports the results for the increase in wage dispersion among high-skill workers between 1964 and 2009 in the model. The table on the right reports the same information for low-skill workers. The other rows for both tables report the results for the increase in wage dispersion for high-skill and low-skill workers, respectively. The second row uses the model without migration costs, the third row without housing, the fourth row without agglomeration and, finally, the fifth row without SBTC.

Table 11: $\Delta\frac{H}{L}$ 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) reports 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) report 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 12: Gross Migration Flows: Correlation Data and Model

| | (1) Corr. Data and Model mean | count |
|---|---|---|
| Migration Rate High-skilled | .5225467 | 38 |
| Migration Rate Low-skilled | .4722791 | 38 |
| Observations | 38 | |

Note: This table reports the correlation between the migration rates generated by the model and the migration rates observed in the data, for both high-skill and low-skill workers over time.

Table 13: Convergence Rates by College Degree 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) |

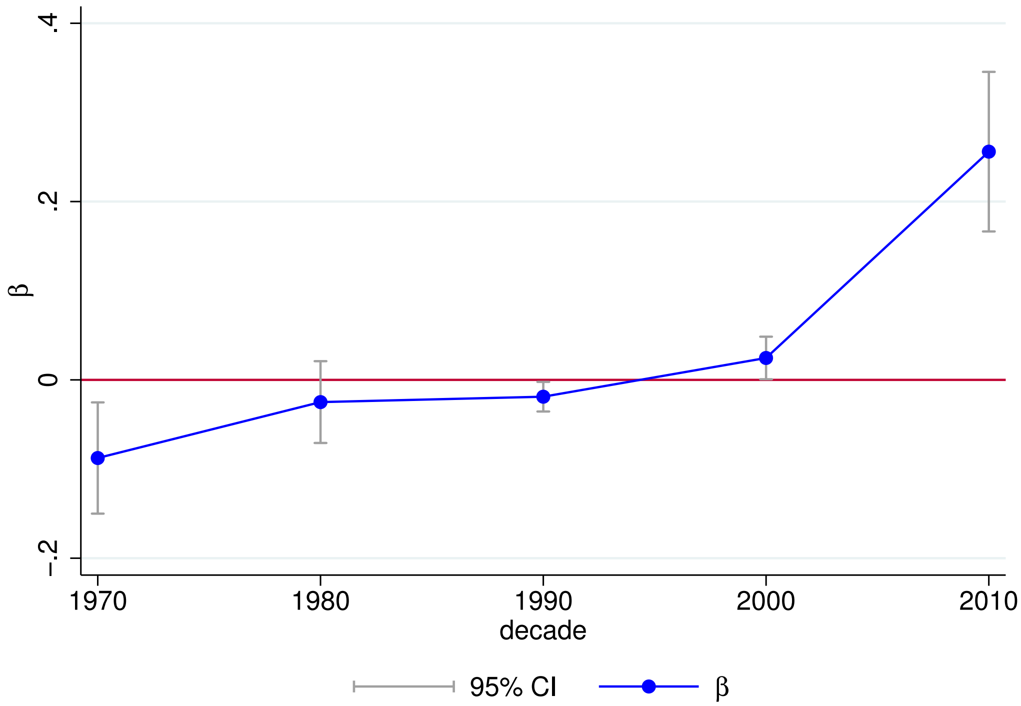
$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Note: The dependent variable in this table is $\Delta w_{jt}$ 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 also for college degree.

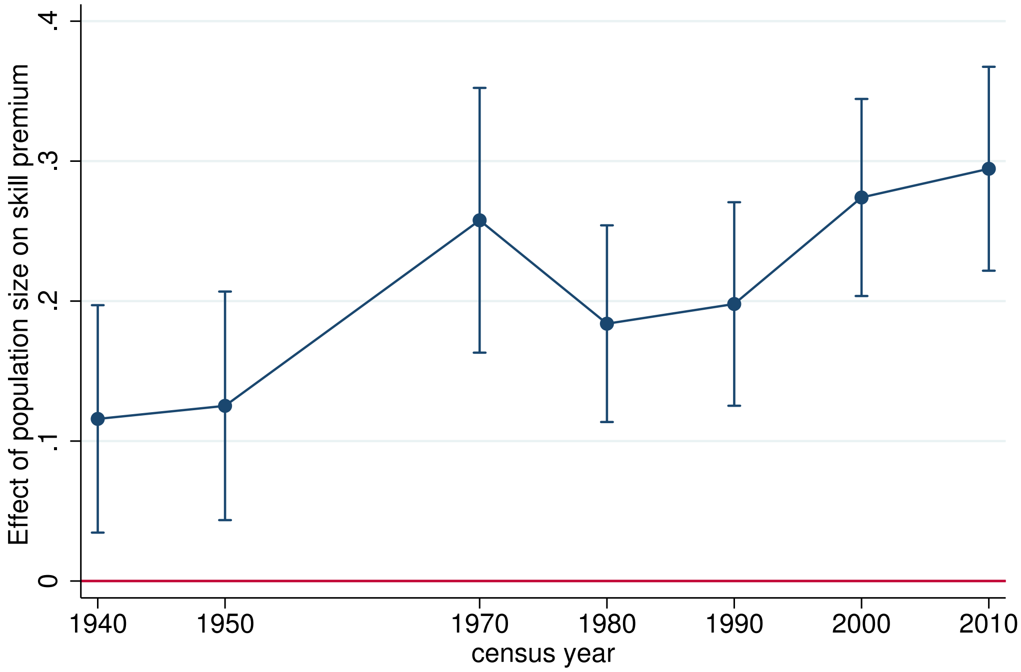
## A.2 Figures

Figure 12: Skill Premium by MSA Education Levels



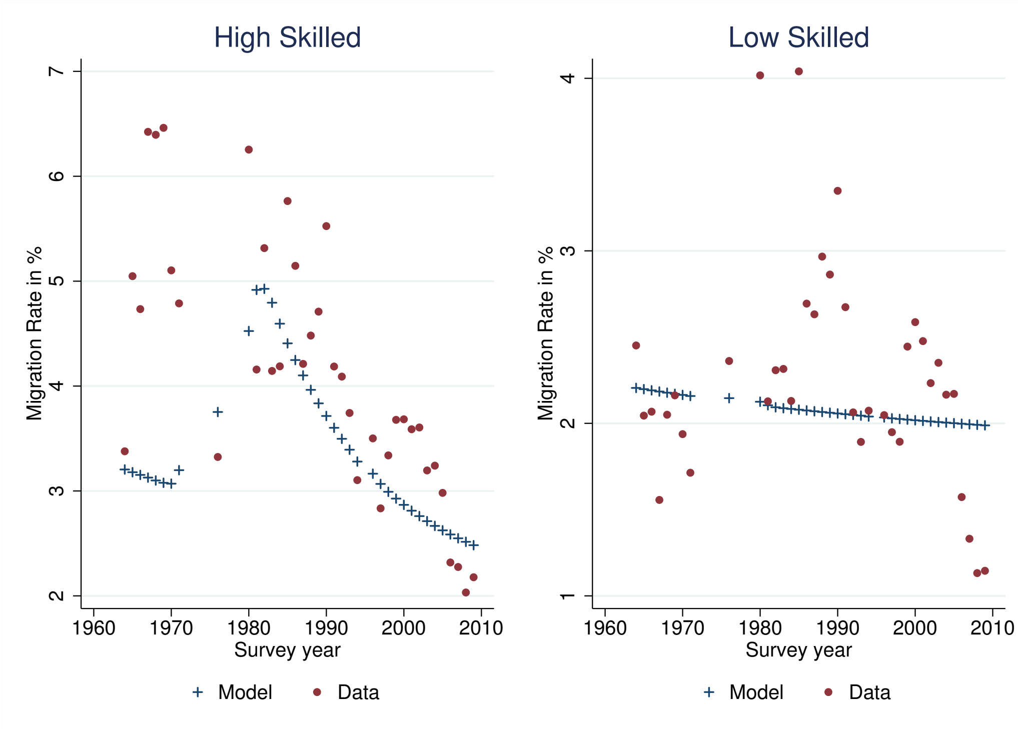

Note: This figure plots the estimate of the coefficient $\beta$ for the regression 1 using CPS data. On the horizontal axis, I have the decades from 1970 to 2010, while on the vertical axis, I have the estimate of coefficient $\beta$ for each decade from 1970 to 2010. Moreover, there is a line starting at 0 on the vertical axis.

Figure 13: Skill Premium by MSA Population Levels



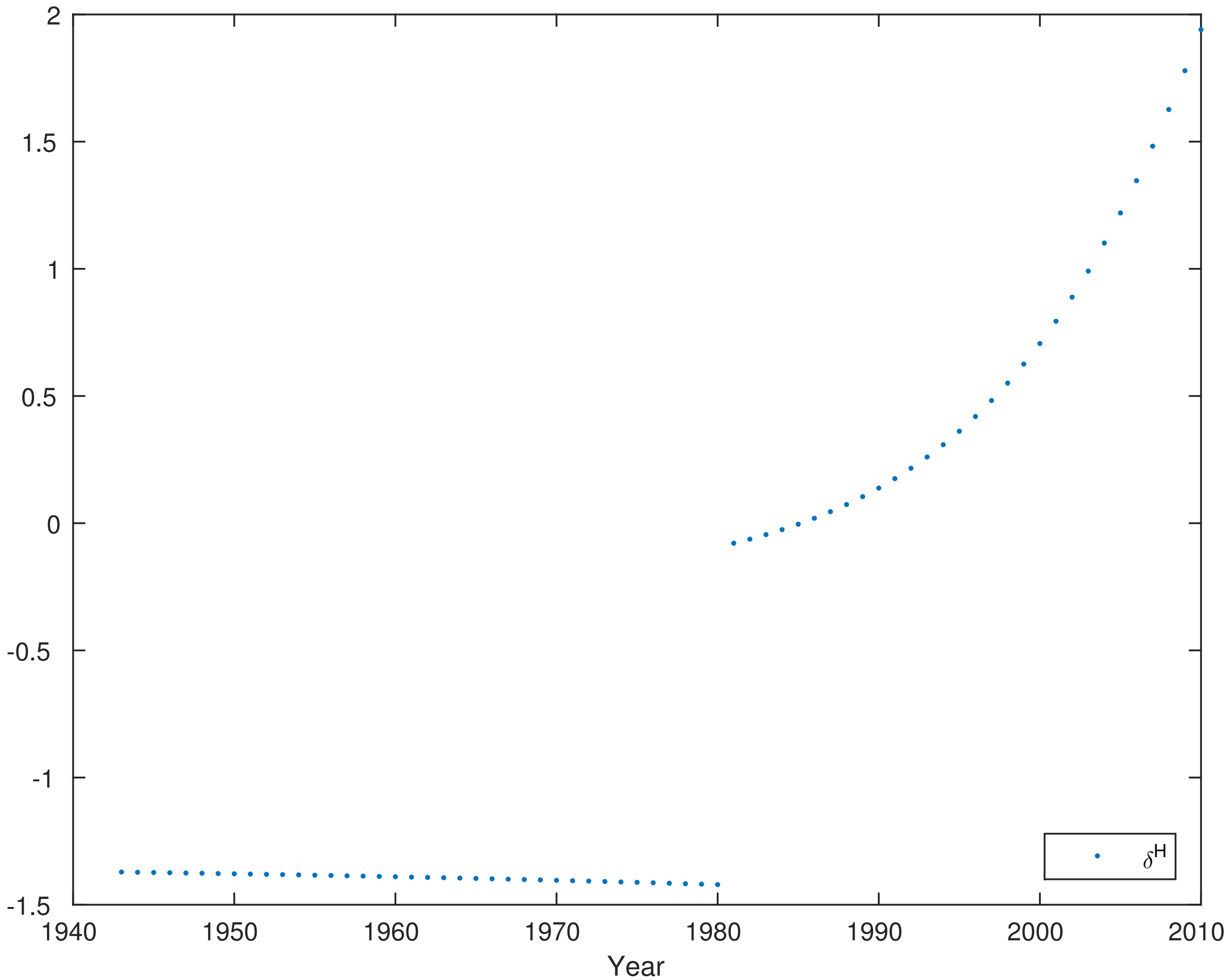

Note: Figure plots the estimate of the coefficient $\beta$ for the regression 1. On the horizontal axis, I have the decades from 1940 to 2010, while, on the vertical axis, I have the estimate of coefficient $\beta$ for each decade from 1940 to 2010. Moreover, there is a line starting at 0 on the vertical axis.

Figure 14: Migration Rate Over Time



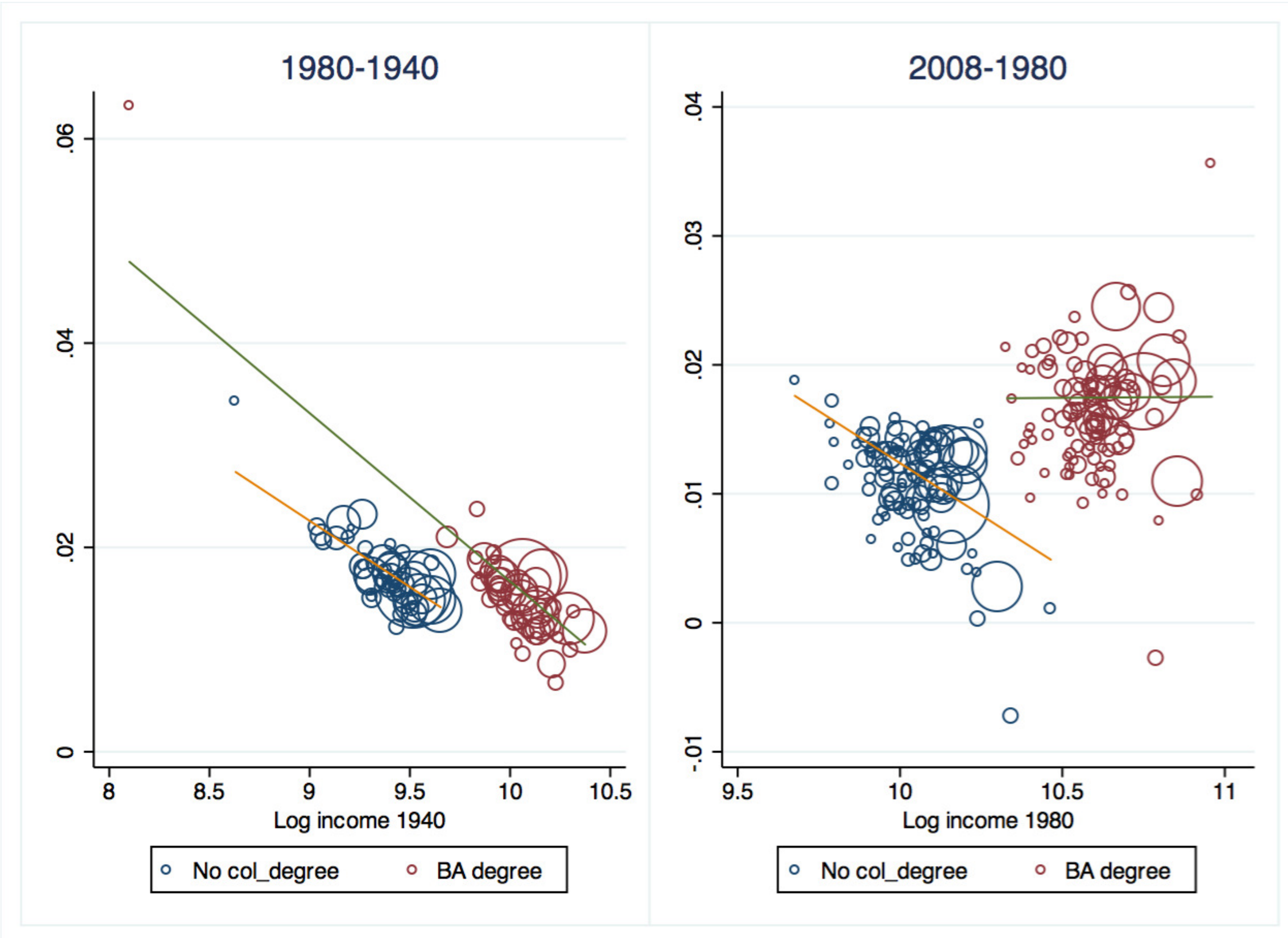

Note: This figure shows the evolution of the migration rate for high-skill and low-skill workers over time for both for model and the data. On the left, I plot the migration rates generated by the model with a cross and those generated by the data with a circle. On the right plot, instead, I plot the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

Figure 15: Sorting Over Time



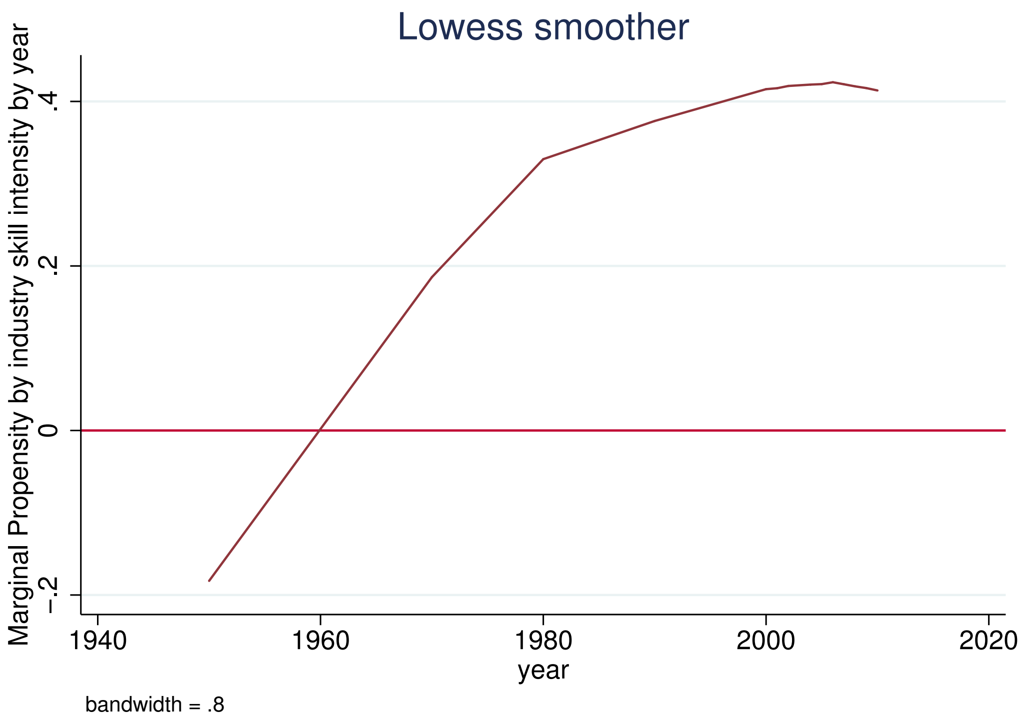

Note: This figure shows the evolution of the estimates of $\delta_t^H$ in equation 35.

Figure 16: Convergence Rates before and after 1980 in low housing elasticity states



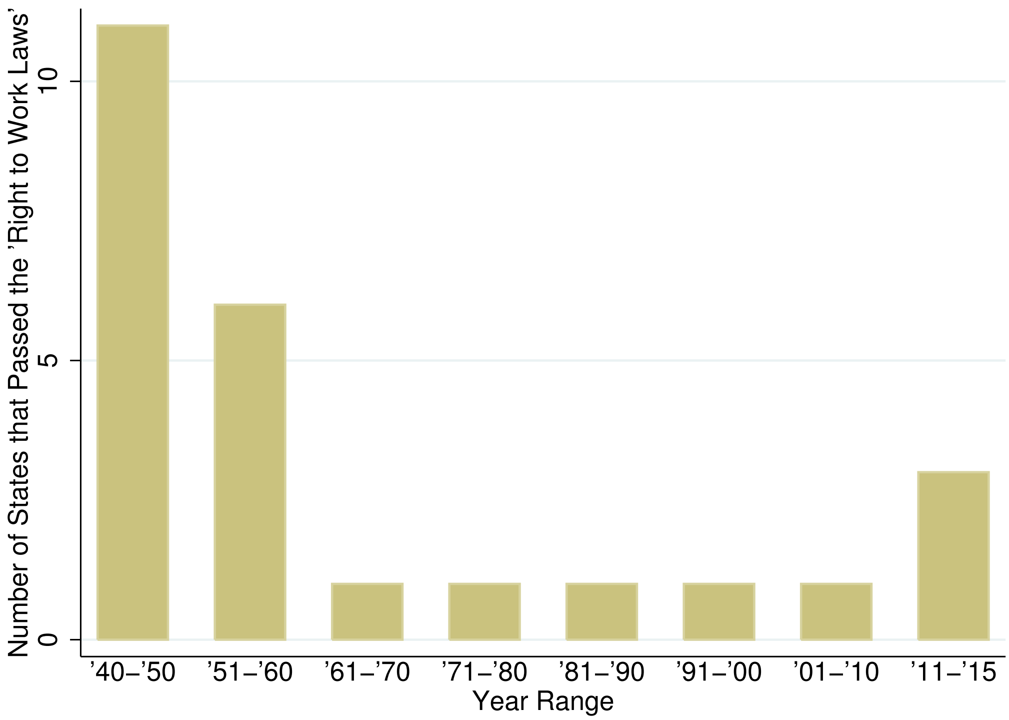

Note: This figure shows two scatter plots of log wages by MSA in the initial year against the annual average growth of wages in the final year by skill type (college degree and non-college degree workers) in cities that are in states with low housing elasticities. In particular, on the left hand side (right hand side), I plot the demeaned log wages in 1840 (1980) by MSA against the annual average growth of wages between 1940 (1980) and 1980 (2010) by skill type (college degree and non-college degree 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 17: Industry Sorting over time




Note: This figure plots the estimated effect of skill concentration at MSA level on skill concentration at industry level. The line is computed using the estimates of the skill ratio at MSA level ($\beta$), using specification 36.

Figure 18: Right To Work Laws




Note: This histogram plots the number of states that passed the "Right to Work Laws" by decade starting with the decade between 1940 and 1950.

## A.3 Data Appendix

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

### A.3.1 Data Description

My two main datasets are the U.S. Census data extracted from IPUMS. I use the 1% sample for 1940, 1% sample for 1950, metropolitan sample 1970, 5% sample 1980, 5% sample 1990, 5% sample 2000. Then, for year the 2010, I use information from 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 this information is also available in the ACS data for 2010. I collect the same information from the CPS dataset. The CPS is a monthly U.S. household survey conducted jointly by the U.S. Census Bureau and the Bureau of Labor Statistics. I use the observation of the month of March. The CPS dataset 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 datasets, 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” dataset, which are available for download at http://online.onetcenter.org/.[32].

### A.3.2 Robustness Checks

Before turning to the robustness tests, I provide the specification for the $\beta$-convergence estimation that I use throughout the paper. The specification is simply

$$\Delta w_{jt} = \beta^o + \beta w_{j\tau} \tag{37}$$

where $j$ is the location unit, $w_{j\tau}$ is the wage at the initial time $\tau$ and $\Delta w_{jt}$ is the annual wage growth between time $t$ and time $\tau$. The estimate of $\beta$ is what I consider as the convergence rate. In most of the specifications, the observations are weighted by the initial size of the locations $j$.

Using the specification above, I run several robustness tests starting with the ones illustrated in figure 1 and in figure 2. I change the unit of analysis, which instead of cities, is counties in figure 19. In figure 19, I plot the estimated convergence rates. In plot A, the estimate uses a 10-year rolling period, while in plot B, for a 20-year rolling period . The convergence rate was negative and statistically significant until 1987 in plot A, while it was negative and statistically significant until 1997 in plot B. Both estimates suggest that the first period in which convergence ceased to be significant was 1978. This fact aligns with the findings of Higgins et al. (2006), which suggests that there was convergence between 1970 and 1990. However, departing from this prior work, I conduct an analysis in which the time period is extended and find that the convergence across counties follows the same patterns as the convergence across cities and states.

As a second robustness check, I show that the rate of convergence stops being significant and robust only if the initial year is after 1980. For this reason, I compute the rolling twenty year wage convergence as shown in figure 20 from 1950 onward. Then, I decompose it by skill group: figure 21 reports results for the high-skill group and 22 for the low-skill. As seen in figure 20, the rolling convergence rate $\beta$ was negative and statistically different from 0 until 1980, but then, it started becoming positive but was still not significant. Finaly, between 1990 and 2010, it became positive and statistically different from 0. But, when I decompose by skill groups, in figure 21, the high-skill workers show the same patterns as the aggregate convergence rate. Instead, in figure 22, the convergence rate for the low-skill group remains negative independently of the time period. It actually becomes even stronger over time.

As a third robustness check, I reproduce figure 1 and 2 using compositionally adjusted wages.

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

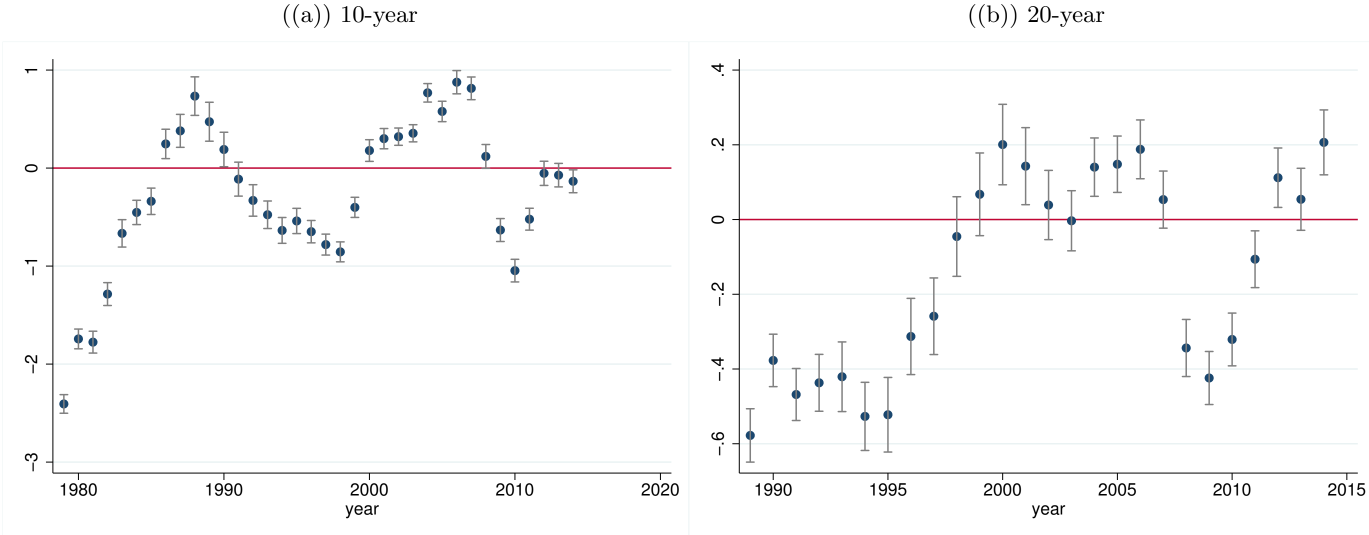
I control whether, after adjusting for observable at individual level (i.e. using compositionally adjusted wages) the convergence rates change. As shown in figure 24, the convergence rates do not change substantially after adjusting for skill composition. Finally, another test is to see whether real wage convergence changes in the same way as nominal wage convergence. The caveat here in looking at real wage convergence is that 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 0 in the low-skill group but becomes positive in the high-skill group.

One reason why the convergence patterns may 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 slow-down in convergence, I estimate the unconditional cities' wage convergence between 1980 and 2010 using only the 127 cities available in 1940-1980. Table 15 shows the convergence rate after 1940 for the reduced sample. The results suggest 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:

$$\Delta w_{jt} = \beta^o + \beta w_{jt-\tau} + \alpha^H \Delta S_{Hjt} + \alpha^L \Delta S_{Ljt} \tag{38}$$

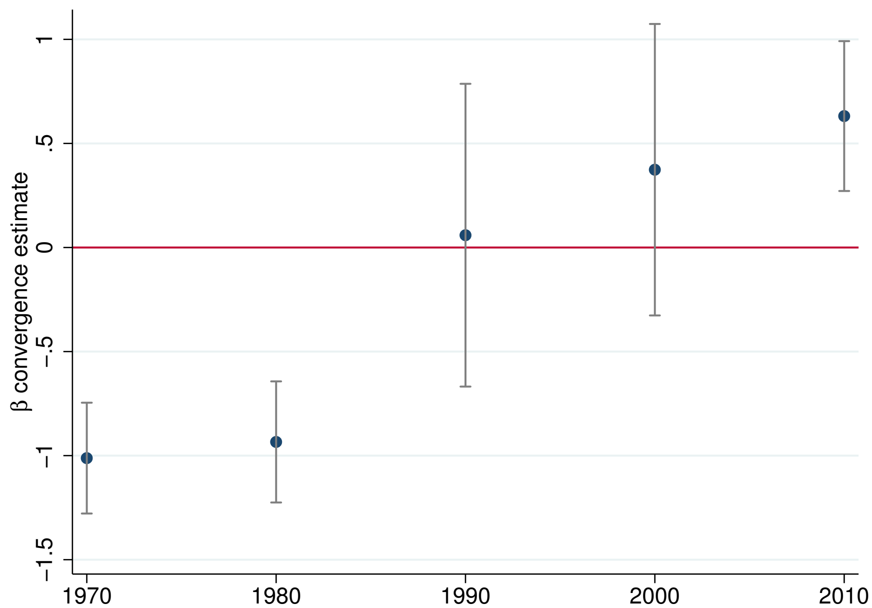
where $t$ is 2010 and $tau$ is 30 years. After controlling for the technology shock, I get conditional convergence $=$ -1.1% a year. This suggests that without taking into account the mechanisms of the model, SBTC affected the decline of wage convergence.

Figure 19: Convergence by county over time



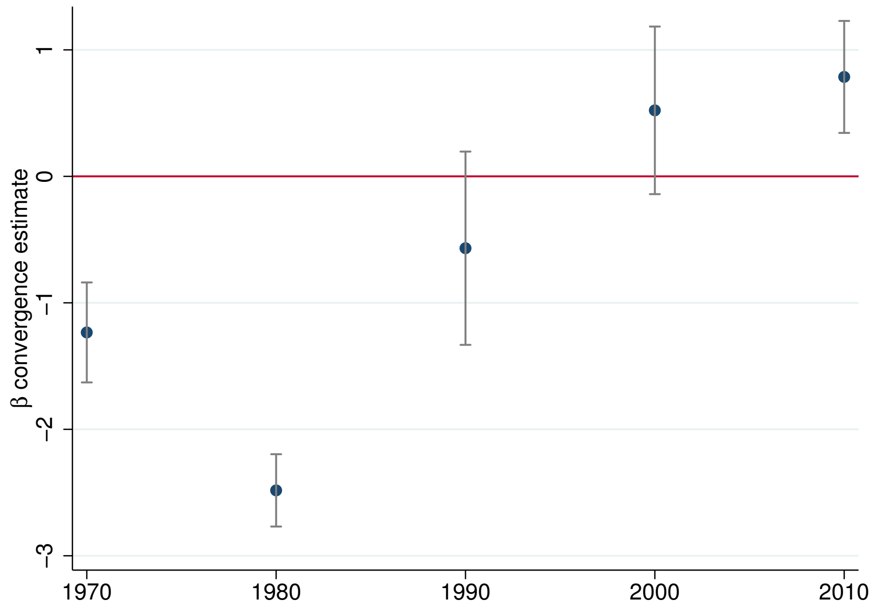

Note: Plot A shows the convergence rate at the county level for a 10-year rolling window starting in 1969. Plot B shows the convergence rate at county level for a 20-year rolling window starting in 1969. Data for this analysis are from the Bureau of Economic Analysis Regional Economics Accounts. In each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

Figure 20: Evolution of $\beta$ convergence rate over time



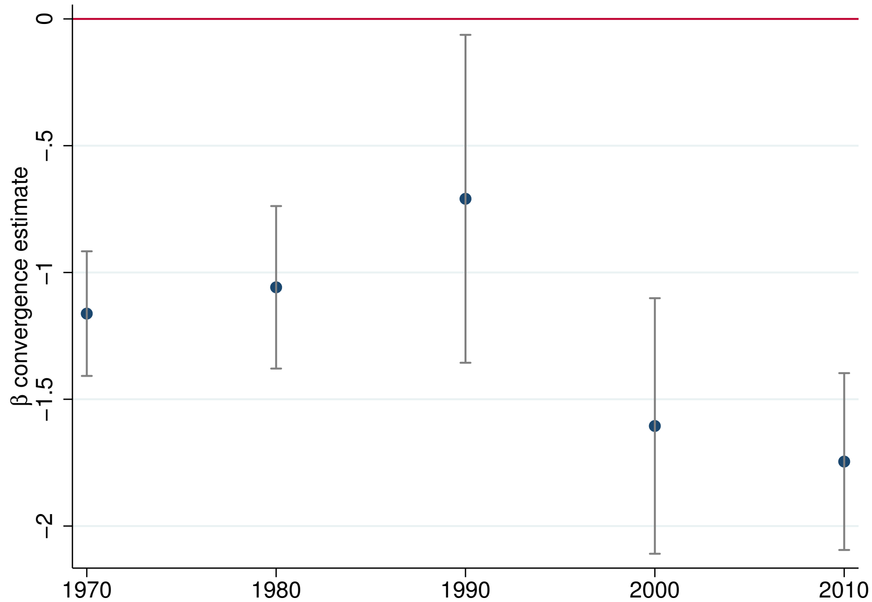

Note: This figure shows the beta coefficient of the regression of the initial wage on the log wage changes using a 20-year rolling window. In each estimate the cities are weighted by their population. On the y-axis the coefficient is reported in percentage terms.

Figure 21: Evolution of $\beta$ convergence rate over time for the high-skill




Note: This figure shows the beta coefficient of the regression of the initial wage on the log wage changes using a 20-year rolling window for high-skill workers only. In each estimate the cities are weighted by their population. On the y-axis the coefficient is reported in percentage terms.

Figure 22: Evolution of $\beta$ convergence rate over time for the low-skill




Note: This figure shows the beta coefficient of the regression of the initial wage on the log wage changes using a 20-year rolling window for low-skill workers only. In each estimate the cities are weighted by their population. On the y-axis the coefficient is reported in percentage terms.

Table 14: Convergence Rates - Restricted Sample

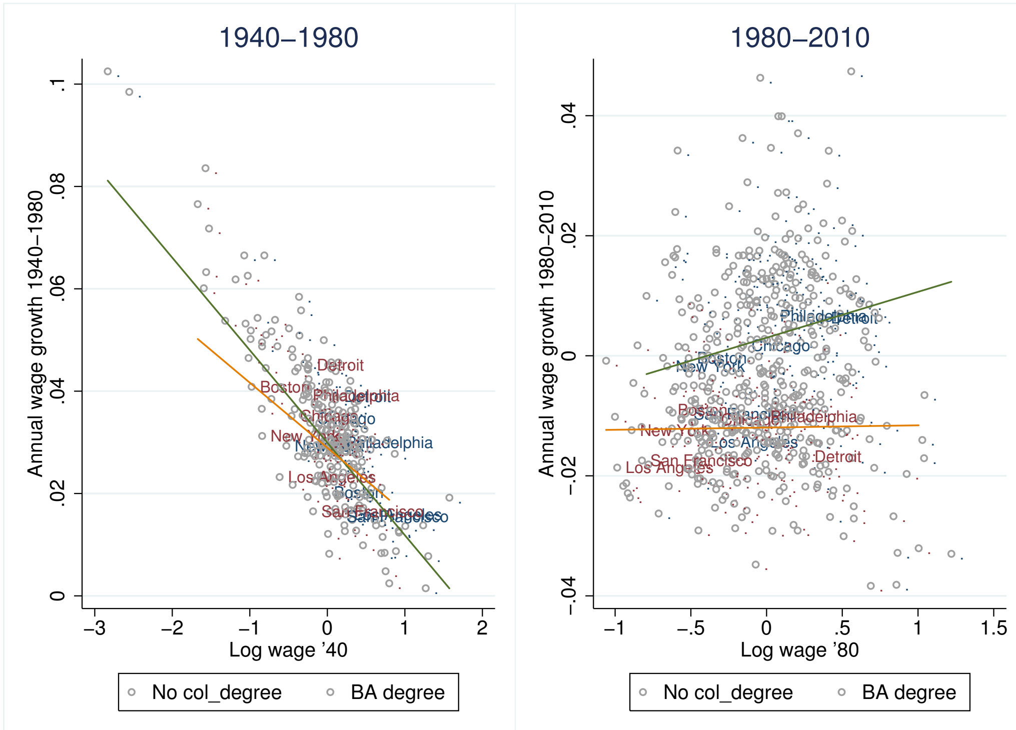
| | (1)<br>$\Delta^{1940-1980}$ | (2)<br>$\Delta_{-}^{80-08}$ |
|---|---|---|
| Log($wage^{1940}$) | -0.0109*** | |
| | (-10.53) | |
| Log($wage^{1980}$) | | -0.00116 |
| | | (-0.25) |
| Constant | -0.0217*** | -0.0147*** |
| | (-137.22) | (-24.45) |

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

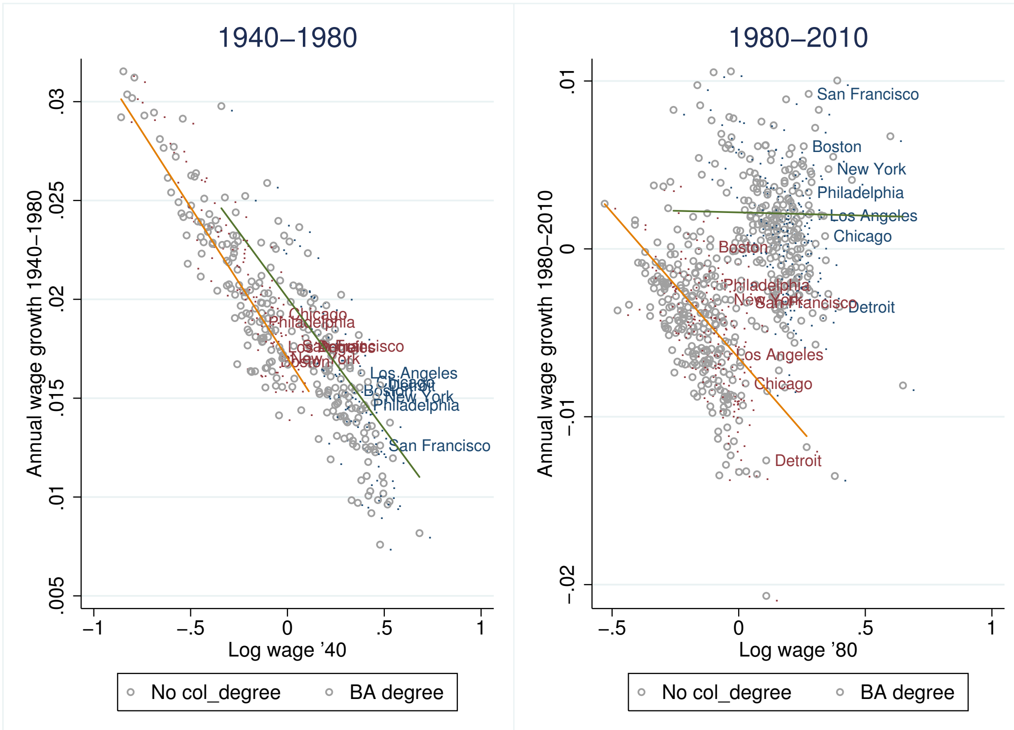
Note: I estimate the $\beta$ convergence rate for the restricted sample with only 127 cities. In column (1), I estimate it for the 1940-1980 time period and in column (2) for the 1980-2010 time period.

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 have been 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 plot of the log wages by MSA in the initial year against the annual average growth of wages in the final year. Wages have been adjusted by individual characteristics, aka 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) $\Delta w^{40-80}\_pw$ | (2) $\Delta w\_pw^{80-10}$ | (3) $\Delta w^{40-80}$ | (4) $\Delta w^{80-10}$ |
|---|---|---|---|---|
| Log($wage^{1940}$) | -0.0137*** | | -0.0214*** | |
| | (-3.38) | | (-13.20) | |
| Log($wage^{1980}$) | | 0.00271 | | -0.00269 |
| | | (1.21) | | (-1.50) |
| Constant | 0.0728*** | 0.0331*** | 0.0749*** | 0.0321*** |
| | (60.16) | (58.69) | (110.74) | (59.42) |

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Panel B

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

This table reports the estimate of the $\beta$-convergence of the OLS regression. In column (1) and (2), the table reports the estimates, respectively, for 1940-1980 and 1980-2010 span using population weighted observations. In columns (3) and (4) the table reports the estimates, respectively, for 1940-1980 and 1980-2010 using unweighted population observations. In Panel A, the table reports the estimates of the $\beta$-convergence for local wages adjusted by the rent in each MSA. In Panel B, the table reports the estimate of the $\beta$-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 reports the estimate of the $\beta$-convergence of the OLS regression. In column (1) and (2) the table reports the estimates, respectively, for "No" college degree and for "Yes" college degree workers for the years 1940-1980. In columns (3) and (4) the table reports the estimates, respectively, for "No" college degree and for "Yes" college degree workers for the years 1980-2010. In Panel A, the table reports the estimates of the $\beta$-convergence by skill for local wages adjusted by the rent in each MSA. In Panel B, the table reports the same estimates as in Panel A but the observations are not weighted by local population. In Panel C, the table reports the estimate of the $\beta$-convergence for compositionally adjusted wages. Panel D reports the same results but the observations are not weighted by MSA population.

Table 17: Convergence Rates and SBTC

| | (1)<br>A | (2)<br>B | (3)<br>C | (4)<br>D | (5)<br>E |
|---|---|---|---|---|---|
| Log hourly wages 1980 | -0.0000389 | -0.00657* | -0.00802* | -0.00912* | -0.0105* |
| | (-0.02) | (-2.59) | (-2.16) | (-2.13) | (-2.30) |
| $\Delta S_{HjH1980}$ | | 0.0160* | 0.0188 | 0.0182 | 0.0201 |
| | | (2.40) | (1.62) | (1.39) | (1.47) |
| $\Delta S_{Lj1980}$ | | 0.0406*** | 0.0220 | 0.0258 | 0.0233 |
| | | (4.44) | (1.12) | (1.22) | (1.07) |
| $\Delta S_{Hj1970}$ | | | 0.0183* | 0.0184* | 0.0200* |
| | | | (2.41) | (2.25) | (2.37) |
| $\Delta S_{Lj1970}$ | | | 0.0342* | 0.0411* | 0.0464* |
| | | | (2.00) | (2.13) | (2.22) |
| $\Delta S_{Hj1950}$ | | | | -0.00162 | -0.00258 |
| | | | | (-0.51) | (-0.78) |
| $\Delta S_{Lj1950}$ | | | | -0.000279 | 0.00213 |
| | | | | (-0.03) | (0.23) |

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Note: This regression shows the coefficient of a regression of wage growth between 1980 and 2010 on the initial wage in 1980, conditioning on the Autor and Dorn measures of SBTC and adding lags for the measure.

### A.3.3 How Do the Estimates Fit the Data?

In this section, I run an experiment using the estimated parameters and the wage equations. Specifically, I test whether the "residuals of the model predict convergence and how much". The idea of evaluating the convergence in the residuals answers the following question: What would the data predict after I structurally condition on the mechanisms of the model?

Once I have the estimates for the labor demand parameters $\{\rho, \gamma^H, \gamma^L, \gamma, \lambda^H, \lambda^L\}$, I can calculate the change in predicted wages $\Delta \hat{w}_{kjt}$, respectively:

$$\Delta \hat{w}_{Ljt} = (1-\hat{\rho})\Delta \ln \hat{Y}_{Tjt} + (\hat{\rho}-1)\Delta \ln L_{jt} + \hat{\gamma_L}\Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) + \hat{\gamma}\Delta \ln\left(H_{jt}+L_{jt}\right)\hat{\lambda}^L \Delta S_{Ljt}$$

$$\Delta \hat{w}_{Hjt} = (1-\hat{\rho})\Delta \ln \hat{Y}_{Tjt} + (\hat{\rho}-1)\Delta \ln H_{jt} + \hat{\gamma_H}\Delta \ln\left(\frac{H_{jt}}{L_{jt}}\right) + \hat{\gamma}\Delta \ln\left(H_{jt}+L_{jt}\right)\hat{\lambda}^H \Delta S_{Hjt}$$

Using the levels of the same equations, we can calculate the predicted wages $\hat{w}_{Hjt}$ and $\hat{w}_{Ljt}$. Therefore, we can calculate the levels and changes in wage residuals, $\Delta\hat{w}_{Kjt,res}$ and $w_{kjt,res}$, by subtracting the predictions above from the data counterpart:

$$\Delta w_{kjt,res} = \Delta w_{kjt} - \Delta\hat{w}_{kjt}$$

$$w_{kjt,res} = w_{kjt} - \hat{w}_{Kjt}$$

Having information on wage changes and wage levels not explained by the model, we can run a convergence regression in the residuals for each skill type $k$ and for their average, $w_{jt,res}$ :

$$\Delta w_{jt,res} = \alpha + \beta w_{j\tau,res} + \epsilon_{jt}$$

and estimate $\beta$. I also estimate $\beta_H$ and $\beta_L$, respectively for the structural residual wages $w_{Ljt}$ and $w_{Hjt}$.

The question behind this experiment is: Once we "remove" the the skill-biased productivity forces, $\Delta S_{Hjt}$ and $\Delta S_{Ljt}$, and the agglomeration forces, $\frac{H_{jt}}{L_{jt}}$ and $H_{jt} + L_{jt}$, how much wage convergence is left unexplained in the residuals?

The results suggest that $\beta$-convergence in the residuals would be 1.75%, a number fairly close to the convergence between 1940 and 1980 reported in table 18. But, when we re-run the convergence in the residuals with $\gamma^H$, $\gamma^L$ and $\gamma$ set to 0 in the equations above, the $\beta$-convergence is estimated to be 1.26% in the period 1980-2010, suggesting that more convergence is left in the residuals. When agglomeration forces are left in the residuals, we observe less convergence, and thus we conclude that agglomeration forces are a strong divergence force.

In column (2) of table 18 , I substitute my $\Delta S_{kjt}$ with a Bartik shock for $H$ and $L$ as in Diamond (2016). Then, I run the same exercise as above, estimating the convergence in the residuals. The estimates of the $\beta$-convergence are much lower if we compare column (1) with column (2). By looking at the analysis, the model estimated with the Bartik shock can generate a lower convergence rate. This may be due to the fact that a Bartik shock does not isolate technology but may instead be generated by anything that affects the local economy. In columns (3) and (4), I decompose the wage convergence in the residuals by skill group. One interesting finding is that isolation of skill-neutral productivity would have induced a higher level of convergence on average for high-skill than low-skill workers. The objective of this exercise is also to show that if instead of using $\Delta S_{kjt}$, I had used a generic Bartik shock, the model would not have matched convergence as well as the $\Delta S_{kjt}$ does. Therefore, while a general Bartik as used in Serrato and Zidar (2014) and in Diamond (2016) would work in the identification of the parameters, it would not work in matching the non-targeted moments in the decline of wage convergence.

Table 18: Model Convergence Predictions

| $\hat{\beta}^{1940-1980}$ | (1) | (2) | (3) | (4) height |
|---|---|---|---|---|
| | (-0.019***) | | | |
| | SBTC | Bartik | $\geq$Coll. Deg. | $<$Coll. Deg. |
| SBTC | -0.0126*** | -0.0033*** | -0.0136*** | -0.0106*** |
| Spillover | -0.0145*** | -0.0012*** | -0.0165*** | -0.0125*** |

## A.4 Other facts on the workers' skills, wages and migration premium

As extra supporting evidence for the importance of skill agglomeration, I propose an agglomeration index $\Phi_{ikt}$ that differs by MSA $j$, time $t$ and by occupation $k$ as well.[33]

**Skill Agglomeration Index** To check the density of high-skill workers in each occupation, I construct it from the data at individuals level and I check how in each region $j$ and in each occupation $k$ the skills are dense and I will compare across occupations in 2 regions (North and South). I can write the index as

$$\Phi_{jkt} = \frac{h_{jkt}}{\sum_{k=1}^{10}(h_{jkt} + l_{jkt})}$$

*Fact: Skill agglomeration effects are correlated positively with individual wages in the country at each point in time. Moreover, the correlation increases over time.*

For the sake of understanding how the concentration of skills in a region has a positive effect on wages, I run a simple regression for different time periods

$$\ln w_{i,m} = \beta \cdot \Phi_{i,m,k} + \sum_{m=1}^{M} D_m \cdot \pi_m + \varepsilon_{i,m}$$

We see that $\Phi_{i,m,k}$ is always positively correlated with the incomes of the agents but it increases over time. The results of the regression show that the coefficient on $\Phi_{i,m,k}$ is positive and statistically significant in all the regressions but its magnitude grows over time. More specifically, it is equal to 0.829 in the 1971-1980 time period while it is more than the double that amount in 2001-2007.

[33] I will use the 10 occupation categories.

Table 19: Impact of skill agglomeration on wages

| | (1) 1971-1980 | (2) 1981-1990 | (3) 1991-2000 | (4) ¿2000 |
|---|---|---|---|---|
| North | 0.0248*** | 0.0978*** | 0.122*** | 0.0701*** |
| | (6.08) | (26.34) | (29.18) | (21.25) |
| South | -0.124*** | -0.0309*** | 0.0141*** | -0.00336 |
| | (-29.73) | (-8.44) | (3.45) | (-1.04) |
| skill_index | 0.829*** | 1.198*** | 1.655*** | 1.768*** |
| | (48.46) | (91.40) | (156.21) | (224.91) |
| _cons | 9.801*** | 9.604*** | 9.511*** | 9.609*** |
| | (2129.00) | (2476.23) | (2358.93) | (3083.85) |
| $N$ | 379826 | 501336 | 503875 | 753471 |

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

*Fact: The employment concentration index is positively correlated with individual wages once I control for regional composition.*

Another relevant analysis would be to study the effect of the employment concentration by sector on wages.

$$\ln w_{i,m} = \beta \cdot \Gamma_{i,m} + \sum_{m=1}^{M} D_m \cdot \pi_m + \varepsilon_{i,m}$$

The results show that the coefficient $\beta$ is always positive and statistically significant except during the period $1971 - 1980$. This suggests that being in an occupation in a country where that occupation is more concentrated is positively correlated with wages. Moreover, between 1982 to 2007 this effect has grown over time. During 1981-1990 the effect was 0.305, while in the 2001-2007 time period, the effect was 1.618, a six-fold increase.

Table 20: Employment agglomeration effect on wages by region

| | (1) 1971-1980 | (2) 1981-1990 | (3) 1991-2000 | (4) ¿2000 |
|---|---|---|---|---|
| North | 0.0323*** | 0.111*** | 0.141*** | 0.0802*** |
| | (7.89) | (29.68) | (33.15) | (23.86) |
| South | -0.134*** | -0.0390*** | 0.0138*** | -0.00358 |
| | (-31.93) | (-10.56) | (3.32) | (-1.09) |
| empl. index | -0.116*** | 0.305*** | 1.055*** | 1.618*** |
| | (-6.08) | (20.70) | (77.08) | (151.10) |
| _cons | 9.996*** | 9.799*** | 9.656*** | 9.665*** |
| | (2042.90) | (2381.46) | (2182.10) | (2780.34) |
| $N$ | 379826 | 501336 | 503875 | 753471 |

$t$ statistics in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

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

**Migration Premium** I define a new variable which I call 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 all the workers who moved within the last year and locals the ones that did not. For the worker to be a migrant, he needs to have changed state in the last year. I compute the average of the compositionally adjusted wages for the workers who changed 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.

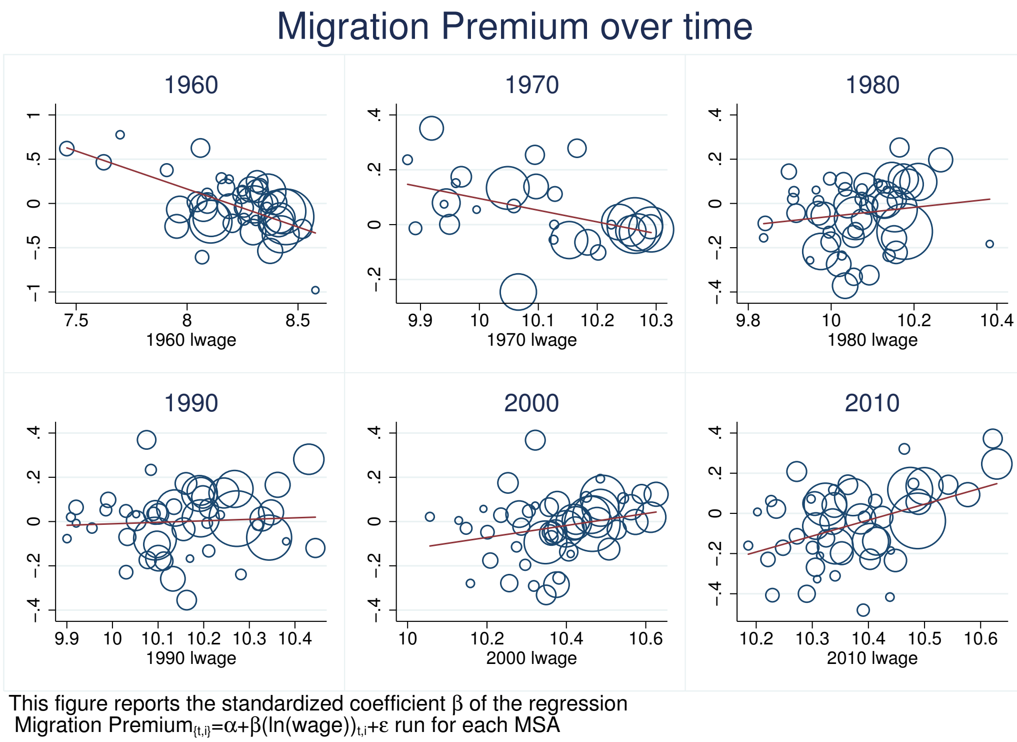
In figure 25, I look at the migration premium over time across states and its evolution. For each of the years in the CPS sample, I run the following specification:

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

I run this specification for all the years of the sample in which the information on migration are 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 was negatively correlated with the wage level of the state while the relationship becomes positive in 1980. I interpret this empirical finding as suggesting that the advantage of migrating until the 1970 was higher in poorer states. While, in the current days it became higher in the richer states.

Figure 25: Migration Premium by State over Time




Note: Figure shows a simple scatter plot of the log of the wages in the state in the first time period $t$ against the migration premium based on our measure of difference between the wages of the migrants and wages of the local 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.

# B Theory Appendix

This appendix supplements the theoretical framework presented in Section 4 in several respects. In section B.1, I describe the algorithm for solving the system of equations and obtain the solution of the model. Section 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. Finally, Section 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 the following 7, which are reported again below for the sake of clarity:

$$W_{Hjt} = (\eta_{Hjt})[\eta_{Ljt} L_{Tjt}^{\rho} + \eta_{Hjt} H_{jt}^{\rho}]^{\frac{1}{\rho}-1} H_{jt}^{\rho-1} \tag{39}$$

$$W_{Ljt} = (\eta_{Ljt})[\eta_{Ljt} L_{Tjt}^{\rho} + \eta_{Hjt} H_{jt}^{\rho}]^{\frac{1}{\rho}-1} L_{Tjt}^{\rho-1} \tag{40}$$

$$W_{Ljt} = P_{Njt} \exp(\xi_{Ljt}) \tag{41}$$

$$\xi_{Ljt} L_{Njt} = (1-\theta) \frac{H_{jt} W_{Hjt} + L_{jt} W_{Ljt}}{P_{Njt}}$$

$$R_{jt}^{\mu} = H_{jt}\left[\bar{O} + (1-\nu)(1-\theta)\frac{W_{Hjt}}{R_{jt}}\right] + L_{jt}\left[\bar{O} + (1-\nu)(1-\theta)\frac{W_{Ljt}}{R_{jt}}\right] \tag{42}$$

From the labor market location decision, labor market clearing becomes

$$H_{jt} = \frac{\exp(\delta_{Hjt}/m_{2H}(j))}{\sum_s^S \exp(\delta_{Hst}/m_{2H}(s))} \tag{43}$$

$$L_{jt} = \frac{\exp(\delta_{Ljt}/m_{2L}(j))}{\sum_s^S \exp(\delta_{Lst}/m_{2L}(s))} \tag{44}$$

where

$$\begin{aligned} \delta_{kjt} = \Bigg[&\theta \log(W_{kjt} - R_{jt}\bar{H}) + \\ &(1-\theta)(1-\nu)[\log((1-\theta)(1-\nu)\frac{W_{kjt}}{R_{jt}} + \bar{O}] + (1-\theta)\nu \log((1-\theta)\nu \frac{W_{kjt} - R_{jt}\bar{O}}{P_{Njt}}] + \\ &+ A_{kjt} + \gamma^{p} \log(H_{jt} + L_{jt})\Bigg] \end{aligned} \tag{45}$$

We end up with a system of 7 equations in 7 unknowns $\{W_{Hjt}, W_{Ljt}, H_{jt}, L_{jt}, L_{Njt}, P_{Njt}\, R_{jt}\, \forall j\}$ for each MSA. Since the analysis includes 240 cities, we have a system of 7x240=1680 equations. I solve this system using an iteration algorithm. The algorithm consists of the following steps.

1. Give the set of parameters $\{\gamma^H, \gamma^L, \gamma, \rho, \gamma_2, \lambda^H, \lambda^L, \theta, \nu, \gamma, \beta_H, \beta_L, \sigma^H, \sigma^L\}$, the initial productivity $\xi_{Lj0}$ and $\xi_{Hj0}$ for all $j$ cities;

2. Start by guessing an allocation of $\{H_{j0}, L_{j0}\}_{j=1}^{J}$ of people across locations;

3. For each location, compute an equilibrium allocation $h_j$, output $Y_{Tj}$ and $Y_{Nj}$, wages $W_{Hj}$ and $W_{Lj}$ and $P_{Nj}$;

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 $\{H_{j0}, L_{j0}\}_{j=1}^{J}$ is 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 start the loop. Continue the procedure until the distance is smaller than the tolerance level $e^{-10}$.

I solve the model for 70 time periods where time $t$ is years. In the first 40 time periods, $S_{Hjt}$ and $S_{Ljt}$ will be set to 0, then, it will be set to the value from the data from S and $\lambda$ will come from the model estimation. Start looking for the equilibrium at time $t = 0$ and give a value for $\xi_{j0}^{H}$ and $\xi_{j0}^{L}$ 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 suggest 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, we 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 high-skill workers, there will be also a higher ratio of computers per worker.

The production function with the old technology $K$ is equal to

$$Y_g = \mathrm{K}^{(1-\alpha)}[aH^{\rho} + (1-a)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{46}$$

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

$$Y_g = \mathrm{PC}^{(1-\alpha)}[bH^{\rho} + (1-b)L^{1-\rho}]^{\frac{\alpha}{\rho}} \tag{47}$$

where $a < b < 1$, which we can think of as 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 will be increasing in $\frac{H}{L}$. Before the availability of the $PC$ technology, locations $j$ that had higher supply of skilled labor also had relatively low-skill wage (because of a congestion effect on skills). Therefore, the return to skill increases most in locations which 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 supply of skill and the return to skill is given by

$$\ln \frac{W_H}{W_L} = \begin{cases} \ln\left[\frac{aH^{\rho-1}}{(1-a)L^{\rho-1}}\right] & if\ \frac{H}{L} \leq \phi^L \\ \ln\left[\frac{a\phi^{L\rho-1}}{(1-a)}\right] = \ln\left[\frac{b\phi^{H\rho-1}}{(1-b)}\right] & if\ \phi^L < \frac{H}{L} < \phi^H \\ \ln\left[\frac{bH^{\rho-1}}{(1-b)L^{\rho-1}}\right] & if\ \frac{H}{L} \geq \phi^H \end{cases} \tag{48}$$

where $\phi^H$ and $\phi^L$ 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 1 implies 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 0. 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, we 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 $\{\rho,\ \gamma^H,\ \gamma^L,\ \gamma\}$ and the data $\{w_{Hjt},\ w_{Ljt},\ L_{Tjt},\ H_{jt},\ L_{Tjt}\}$. In order to make this transformation, we follow Diamond (2016) by taking the ratio of high-skill wages to low-skill wages in location $j$:

$$\frac{w_{Hjt}}{w_{Ljt}} = \frac{\xi_{Hjt} N_{jt}^{1-\rho} H_{jt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (H_{jt} + L_{jt})^{\gamma}}{\xi_{Ljt} N_{jt}^{1-\rho} l_{Tjt}^{\rho-1} \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (H_{jt} + L_{jt})^{\gamma}} \implies$$

I use a change of variable, defining high-skill and low-skill prodictivities as

$$\xi_{Hjt} = \theta(1 - \lambda_{jt})$$

$$\xi_{Ljt} = \theta(\lambda_{jt})$$

This implies that the skill premium can be written as:

$$\frac{w_{Hjt}}{w_{Ljt}} = \frac{\theta_{jt}^{\frac{1}{\alpha}}(1-\lambda_{jt})N_{jt}^{1-\rho}H_{jt}^{\rho-1}\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H}(H_{jt}+L_{jt})^{\gamma}}{\theta_{jt}^{\frac{1}{\alpha}}\lambda_{jt}N_{jt}^{1-\rho}l_{Tjt}^{\rho-1}\left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L}(H_{jt}+L_{jt})^{\gamma}} \implies$$

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

$$w_{Hjt}l_{Tjt}^{\rho-1}\lambda_{jt} = H_{jt}^{\gamma^H+\rho-1-\gamma^L}L_{jt}^{-\gamma^H+\gamma^L}w_{Ljt} - \lambda_{jt}H_{jt}^{\gamma^H+\rho-1-\gamma^L}L_{jt}^{-\gamma^H+\gamma^L}w_{Ljt} \implies$$

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

$$\implies Y_{Tjt}(\rho,\gamma^H,\gamma^L,\gamma) = \left[\frac{H_{jt}^{\gamma^H+\rho-1}L_{jt}^{-\gamma^H}w_{Ljt}l_{Tjt}^{\rho}(H_{jt}+L_{jt})^{\gamma}}{w_{Hjt}l_{Tjt}^{\rho-1}+w_{Ljt}H_{jt}^{\gamma^H+\rho-1-\gamma^L}L_{jt}^{\gamma^H-\gamma^L}} + \frac{H_{jt}^{\gamma^H+\rho}L_{jt}^{-\gamma^H}w_{Hjt}l_{Tjt}^{\rho-1}(H_{jt}+L_{jt})^{\gamma}}{w_{Hjt}l_{Tjt}^{\rho-1}+w_{Ljt}H_{jt}^{\gamma^H+\rho-1-\gamma^L}L_{jt}^{\gamma^H-\gamma^L}}\right]^{\frac{1}{\rho}}$$

This formulation of $Y_{Tjt}$ will be used in the estimation since it does not include the productivity terms $S_H$, $S_L$, $\xi_H$ and $\xi_L$.