

SKILL-BIASED TECHNICAL CHANGE AND REGIONAL CONVERGENCE*

Elisa Giannone

CREI

October 2026

Abstract

In the last century, the U.S. witnessed the great reversal of regional convergence: between 1940 and 1980, the wage gap between poorer and richer U.S. cities shrank at an annual rate of 1.27% but stopped shrinking after 1980. This paper documents that this change was driven mostly by highly skilled workers. Based on this evidence, I build a dynamic spatial equilibrium model with regional convergence and divergence forces. The model highlights the role of technology: convergence forces enter through spatial technology diffusion and divergence forces through a Skill-Biased Technical Change shock transmitted through local agglomeration and sorting. Two main quantitative results emerge: (i) spatial technology diffusion is consistent with the patterns of regional convergence between 1940 and 1980, while after 1980 local agglomeration is the main channel through which differences in Skill-Biased Technical Change exposure contribute to the decline in regional convergence, especially for highly skilled workers; and (ii) the same mechanism generates a decline in secular migration and the “Great Divergence of Skills.”

*I am indebted to Erik Hurst, Ralph Ossa, Nancy Stokey, and Jonathan Dingel for their invaluable guidance and support throughout this project. I also want to thank Big Banterngans, Gady Barlevi, Robert Barsky, Paco Buera, Lorenzo Caliendo, Jeff Campbell, Rebecca Diamond, Cristian Dagnino, David Dorn, Jason Faberman, Pablo Fajgelbaum, Cecilie Gaubert, Alejandro Justiniano, Munseob Lee, Kerstin Holzheu, Greg Kaplan, David Nagy, Nuno Paixao, Steve Redding, Esteban Rossi-Hansberg, Jaume Ventura and many other seminar participants. I thank those at the University of Chicago, LSE Geography, Marco Fanno First Meeting, the Econometric Society World Congress in Philadelphia, Chicago Fed, Junior Scholar Conference at the Minneapolis Fed, UIUC, FSU, St. Louis Fed, Toronto Rotman, Princeton, Georgetown, UVA-Darden, NYU, Penn State, IIES, Rochester, USC-Marshall, USC, UCLA, SOLE, NBER Macro across Space and Time, SED-Edinburgh, Queen’s, Maryland, CEPR-EBRD and Barcelona GSE Summer Forum, NBER SI Urban, BU, Minneapolis Fed and Berkeley. I acknowledge hospitality from the Chicago Fed where part of this work was completed. Special thanks to Che-Lin Su, who taught me a great deal about numerical methods. I acknowledge financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S). I also acknowledge financial support through the Ramón y Cajal programme (grant RYC2024-049489-I). Any errors are my own.

1 Introduction

A growing literature emphasizes the role of spatial technology diffusion in regional development, whereby productivity gains in more advanced locations propagate to less advanced ones. In parallel, another strand of the literature shows that technological progress has disproportionately benefited highly skilled workers since around 1980, widening the productivity gap with less skilled workers. This phenomenon, known as Skill-Biased Technical Change (SBTC), has led to a substantial increase in the skill premium.^{1,2} This paper adds a third observation about the technological process. I argue that not all highly skilled workers benefited equally from SBTC. Those in cities with a higher concentration of skills and population had a larger increase in wages through agglomeration forces. Combining these three observations, this paper rationalizes one of the most salient features of regional growth in the last century: the great reversal of regional convergence. The same mechanism generates a decline in secular migration across regions. Secular migration describes changes in the geographic distribution of each skill group, measured as 100 times the sum across cities of absolute annual changes in their shares of national employment within that group. Gross inter-MSA mover rates provide a separate measure of individual mobility.

It has been noticed that between 1940 and 1980, wages grew 1.27% per year faster in poorer U.S. cities than in richer ones.³ Regional convergence ended in 1980, and from then to 2010, wages grew at similar rates in cities with different income levels. Figure 1 plots the annual demeaned average wage growth against its initial demeaned wage level in logs. The β -convergence rate, given by the slope of the trend line, is -0.0127 between 1940 and 1980 and -0.0012 between 1980 and 2010; the latter estimate is statistically insignificant.⁴

What has not been noticed up to now is that the end of regional convergence masks a crucial heterogeneity: the end of convergence occurred only for highly skilled workers. This is the central empirical contribution of this paper. As shown in Figure 2, before 1980, the regional convergence rate was the same for both highly skilled and less skilled workers. However, since 1980, the wages of less skilled workers have continued to converge at about 1.79%

¹I define highly skilled workers as those with at least a college degree and less skilled workers as those with less than a college degree.

²Comin, Dmitriev and Rossi-Hansberg (2012), Desmet and Rossi-Hansberg (2014) and Eckert and Peters (2022), among others, show how technology diffusion across space pushes regional convergence. Katz and Murphy (1992), Card and DiNardo (2002), Levy and Murnane (1992) and Bound and Johnson (1992), to cite a few, show how SBTC increased the skill premium. The skill premium is defined as the difference between the log wages of the highly skilled and the less skilled workers.

³Throughout the paper, I use “cities” to refer to “Metropolitan Statistical Areas”, the geographical unit used in this paper. A formal definition is provided in Section 2.

⁴Berry and Glaeser (2005) are the first to point to this decline in convergence across cities after 1980. Ganong and Shoag (2017) show a similar reduction in convergence for income per capita across U.S. states after 1980.

Figure 1: Regional Convergence across Cities before and after 1980



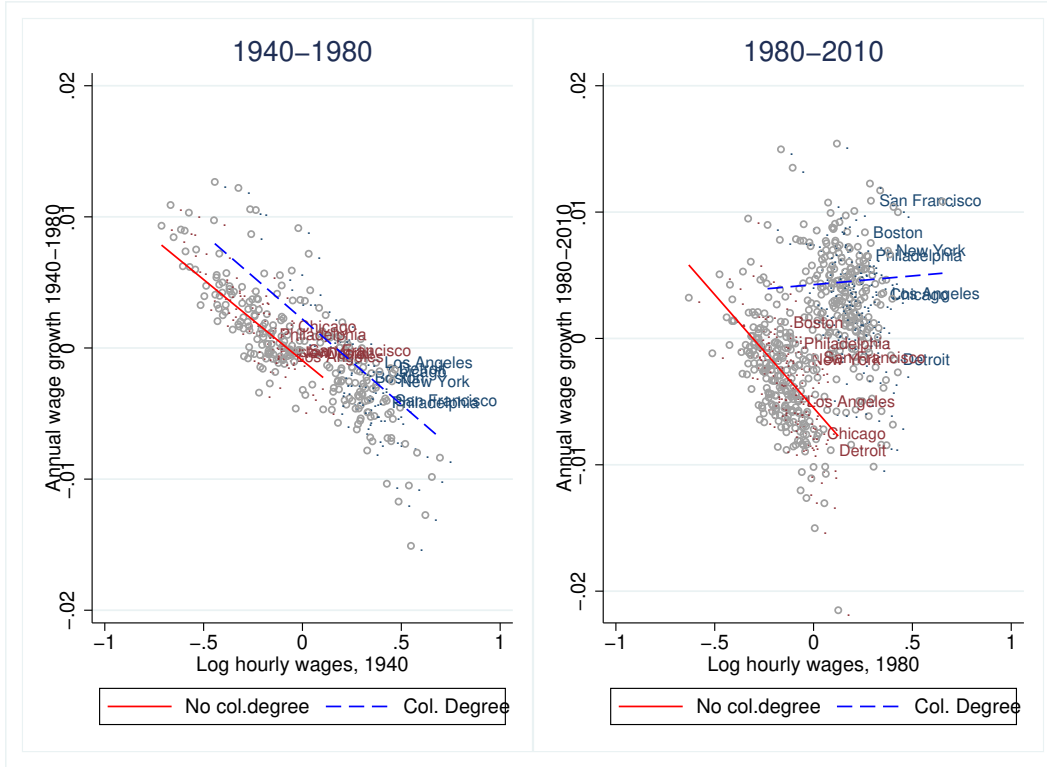
Note: Figure 1 plots each city's (demeaned) annual average wage growth against its (demeaned) initial wage level. The left side depicts 1940–1980; the right side depicts 1980–2010. The red line depicts a weighted least-squares bivariate regression, with each MSA weighted by its population in the initial year.

annually, while the regional convergence rate for highly skilled workers has been essentially zero. Thus, any account of the end of convergence must distinguish between skill groups. This, together with two other empirical regularities regarding local skill premium and selective migration, supports the claim that demand forces, such as agglomeration forces, pushed highly skilled workers toward already highly skilled places, thereby raising their wages.⁵ Overall, performance differences between highly skilled and less skilled workers played a crucial role in the cessation of regional convergence and are also consistent with the idea that demand grew relatively more than supply in cities with a greater initial concentration of highly skilled workers, benefiting from higher wage growth.

Motivated by this evidence, the paper makes two additional contributions: (i) it develops a quantitative dynamic spatial equilibrium model in which technological progress rationalizes regional convergence and its reversal only for highly skilled workers; and (ii) it shows how the

⁵I define the college ratio (or skill ratio) as the ratio of college-educated workers to less-than-college-educated workers. The skill premium is defined as the difference between the log wages of college-educated workers and those of less-than-college-educated workers.

Figure 2: Regional Convergence by Skill Group across Cities before and after 1980



Note: Figure 2 plots each MSA’s annual average wage growth (demeaned) against its (demeaned) initial wage level by skill type (highly skilled and less skilled workers). The left panel depicts 1940–1980; the right depicts 1980–2010. Within each skill group, the regressions weight each MSA by its initial population in that skill group. The solid red and dashed blue lines in each graph depict weighted least-squares bivariate regressions for less skilled and highly skilled workers, respectively.

same mechanism reshapes the geographic allocation of workers across cities.⁶ The model’s main ingredients are spatial technology diffusion, which promotes convergence, and SBTC and local agglomeration forces, which promote divergence. Multiple industries, housing, amenities, and costly trade and migration provide the general-equilibrium environment needed for the inversion and the counterfactuals. Labor productivity varies across local labor markets through endogenous agglomeration forces that depend on city population and skill concentration, a national SBTC component, and a city-specific technology process. Given this set of ingredients, the novelty of the model is to combine two streams in the literature. First, I follow the literature pioneered by [Desmet and Rossi-Hansberg \(2014\)](#) and [Desmet, Nagy and Rossi-Hansberg \(2018\)](#) and introduce spatial technology diffusion, which makes the model dynamic.

⁶The same framework organizes the “Great Convergence” and “Great Divergence” of skills before and after 1980. Following the definition of “Great Divergence” coined by [Moretti \(2012\)](#), I define the “Great Convergence” of skills as the period between 1940 and 1980, when the relationship between the initial college ratio and subsequent college-ratio growth was negative.

Second, in the spirit of [Rosen \(1979\)](#), [Roback \(1982\)](#), [Autor and Dorn \(2013\)](#), and [Diamond \(2016\)](#), I endogenize household location choices across heterogeneous cities and skills. It is crucial to note that accounting for the reversal in wage and employment trends and the associated changes in spatial allocation by skill group requires combining these two streams within the paper’s single quantitative model.⁷ Overall, allowing for heterogeneity in skills and spatial diffusion of technology over time produces a suitable framework for studying why regional convergence reversed and how the model’s mechanisms contribute to that reversal.

The mechanism works as follows. Spatial technology diffusion promotes wage convergence, but after 1980 local agglomeration becomes the main mechanism amplifying differences in SBTC exposure: heterogeneous SBTC exposure is capitalized into larger wage gains in already skill-abundant cities. Sorting reinforces those gains, while rising rents and congestion reduce the incentive for continued migration. I apply the model to the data in the third part of the paper. A static GMM system estimates the agglomeration and utility parameters using local SBTC exposure and Bartik-style instruments. I then estimate high- and low-skill productivity persistence separately on the positive city–sector productivity panel.

A counterfactual analysis points to a joint SBTC–agglomeration propagation mechanism. SBTC shifts local skill demand, while agglomeration and sorting propagate heterogeneous exposure across cities. Agglomeration accounts for about 32 percent of the model’s post-1980 shift toward high-skill divergence. Shutting down SBTC as well increases the agglomeration-only effect by about 14 percent, so the joint shutdown accounts for about 36 percent of the shift. Together, these forces also account for 94 percent of the model’s decline in secular migration among highly skilled workers and more than 100 percent of the decline among less skilled workers, whose decline becomes a small increase when the mechanisms are removed.

Among the other counterfactuals, none emerges as a primary explanation for the high-skill reversal. Human-capital growth and housing affect its magnitude, trade with the Rest of the World has little effect, and migration frictions dampen rather than generate divergence. The results remain stable across alternative SBTC measures. Physical capital instead produces high-skill convergence, while eliminating diffusion generates extreme paths. These exercises reinforce agglomeration and its interaction with SBTC as the central quantitative mechanism.

Related Literature. This paper speaks to four strands of the literature. It contributes to the literature on dynamic quantitative spatial equilibrium models (e.g., [Desmet and Rossi-Hansberg 2014](#), [Desmet, Nagy and Rossi-Hansberg 2018](#), [Caliendo, Dvorkin and Parro 2019](#), [Lyon and Waugh 2019](#) and [Nagy 2023](#)) by incorporating both convergence and divergence

⁷Appendix Section [D.8](#) reports the deliberately extreme no-diffusion benchmark. Removing diffusion over the whole horizon materially changes both skill-specific paths; it is therefore a sensitivity exercise, not an untargeted validation of the transition law.

forces within a unified framework. This framework features heterogeneous skills, industries, housing, and agglomeration, allowing for an analysis of regional and aggregate long-run growth and inequality. The closest work is the seminal paper by [Desmet, Nagy and Rossi-Hansberg \(2018\)](#), from which this model borrows the spatial diffusion of technology. Unlike [Desmet, Nagy and Rossi-Hansberg \(2018\)](#), this model endogenously incorporates rich regional divergence forces and skill heterogeneity, providing a richer understanding of regional convergence, its reversal, and changes in secular migration.

Furthermore, this paper complements the literature on increasing U.S. spatial dispersion and the “Great Divergence” of skills (e.g., [Berry and Glaeser 2005](#), [Moretti 2012](#), [Hsieh et al. 2019](#), [Eeckhout, Pinheiro and Schmidheiny 2014](#), [Hsieh and Moretti 2019](#), [Diamond 2016](#), [Ganong and Shoag 2017](#), [Baum-Snow, Freedman and Pavan 2018](#), [Fajgelbaum and Gaubert 2020](#), [Eckert 2019](#) and [Rossi-Hansberg, Sarte and Schwartzman 2026](#)). Relative to [Berry and Glaeser \(2005\)](#), who focus on the divergence of local human-capital composition across cities, I separate changes in cities’ skill composition from changes in wages within skill groups. This distinction is central to my analysis of how SBTC and agglomeration transform pre-existing spatial skill differences into divergent local wage paths. It adds to this literature empirically by showing that regional convergence only stopped for highly skilled workers and that a period of “Great Convergence” preceded the divergence. The model incorporates spatial technology diffusion and uses counterfactual experiments to quantify how convergence and divergence forces shape regional wages and secular migration. Unlike the “skill-biased amenities” explanation in [Diamond \(2016\)](#), this work emphasizes skill-biased technological progress and spatial diffusion.

This paper also connects directly to the literature using regional variation in technological shocks to understand their consequences on wage inequality, such as [Autor and Dorn \(2013\)](#). I build on this line of work by studying how technological change may have transformed regional convergence across U.S. cities into regional divergence led by highly skilled workers. The model allows for the interaction between national productivity shocks and local endogenous migration forces that can rationalize the local incidence of the technology shock first introduced by [Autor and Dorn \(2013\)](#).

Additionally, this paper complements long-standing literature on regional convergence across countries and states inspired by the seminal works of [Baumol \(1986\)](#), [Barro and Sala-i Martin \(1992\)](#), and [Barro and Sala-i Martin \(1997\)](#) and subsequently developed by [Bernard and Jones \(1996\)](#), [Caselli and Coleman \(2001\)](#), [Gennaioli et al. \(2014\)](#) and [Comin and Mestieri \(2018\)](#). My paper complements this literature in several dimensions. First, it develops a quantitative model of U.S. regional economies that connects directly to the data. Second, I propose a framework with both convergence and divergence built in and use it to organize

the observed spatial paths and their counterfactual decomposition.

The remainder of the paper is organized as follows. Section 2 describes the data and the empirical analysis. Section 3 proposes a theoretical framework. In Section 4, I estimate the core parameters and describe the inversion. In Section 4.1, I report the model solution and spatial moments, and in Section 5 I conduct the frozen-wedge counterfactual analysis. Section 6 reports robustness and material sensitivity exercises. Section 7 concludes.

2 Data and Empirical Regularities

This paper’s main new empirical finding is that regional convergence stopped for highly skilled workers but continued for less skilled workers. This fact motivates the exploration of the skill premium and migration patterns over the last 70 years. I also document convergence in skill ratios across MSAs before 1980, followed by divergence after 1980. To the best of my knowledge, the facts in Sections 2.2, 2.3, and 2.4 are new to the literature, while the fact in Section 2.5 reproduces the previously documented “Great Divergence” and introduces the “Great Convergence.”

2.1 Data

The analysis draws on the Census Integrated Public Use Microdata Samples (IPUMS) for each census decade from 1940 to 2000, and the American Community Survey (ACS) for 2010 (Ruggles et al. 2015).⁸ For the main migration analysis, I use the 1950 Census and the 2006 and 2008 ACS, which provide comparable one-year inter-MSA migration measures. A complementary comparison using the March Current Population Survey (CPS) appears in Appendix Figure A.1 and Table A.5. More details about the data and the definitions of the variables are in Appendix Section A. Appendix Table A.1 summarizes the notation used in the empirical analysis.

2.2 The End of Regional Convergence for Highly Skilled Workers

The paper’s central empirical finding is the difference in convergence trends documented in Figure 2 and in this section: β^H is negative before 1980 but approximately zero after 1980, while β^L remains negative. Figure 2 shows that the cross-MSA wage convergence rates

⁸The decennial IPUMS Census samples from 1940 through 1960 are 1% samples; the 1970 metropolitan Census sample and the 2010 ACS sample each include about 1% of the population; and the decennial Census samples from 1980 through 2000 each include 5% of the population. In 1950, however, some key income variables are available only for the sample-line person in each selected household, so the effective person-level coverage for those variables is closer to 1 in 330 (about 0.3%) before additional sample restrictions.

between 1940 and 1980 were the same for highly skilled and less skilled workers. The patterns diverge sharply after 1980. Between 1980 and 2010, the regional wage convergence rate was observed only among less skilled workers, not among highly skilled workers. To illustrate these patterns, I estimate the standard convergence specification introduced by [Baumol \(1986\)](#):

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

where k is the skill group, highly skilled H or less skilled L ; j is the MSA; and t is the final year of the analysis and τ is the initial year. $w_{kj\tau}$ is the log hourly wage by skill group k in MSA j at time τ . The dependent variable is the annual average wage growth of log hourly wages between τ and t . The regression for $k = H$ is weighted by $H_{j\tau}$, and the regression for $k = L$ by $L_{j\tau}$, the corresponding skill-group populations in the initial year τ . If the estimate $\hat{\beta}^k$ is negative and statistically significant, wages exhibit regional β -convergence, and $\hat{\beta}^k$ measures the annual convergence rate. If $\hat{\beta}^k$ is positive and statistically significant, wages exhibit regional divergence. In [Figure 2](#), I plot the observations at the MSA level by skill group k and then the fitted line, where the β^H and β^L convergence rates are the slopes of the lines. The solid red line represents the β -convergence for L , and the blue dashed line indicates the β -convergence for H . Each circle is an observation by MSA and skill group. I label the ten largest U.S. MSAs in red for less skilled and in blue for highly skilled workers.

Between 1940 and 1980, there was no difference between cross-MSA wage convergence rates, β^H and β^L . Between 1980 and 2010, the convergence rate β^L was still negative and statistically significant, but β^H was not. This difference means that the end of convergence was driven primarily by the breakdown in the convergence of wages among highly skilled workers, since the wages of less skilled workers still converged across MSAs. In Panel B of [Appendix Table A.2](#), I report the estimates of β^L and β^H for the two different periods, both for population-weighted and non-population-weighted regressions. For the population-weighted regressions, β^L and β^H are -1.24 and -1.30 annual percentage points, respectively, in 1940–1980. Both estimates are statistically significant. In 1980–2010, β^L is -1.79 annual percentage points and statistically significant, whereas β^H is approximately zero and statistically insignificant. I examine the robustness of this difference between skill groups in [Online Appendix Section OA.1](#). In particular, I show that the result is not driven by changing MSA composition over time by re-estimating [Fact 2](#) on a fixed pre-1980 MSA sample (Panel A of [Online Appendix Table OA.2](#)); I also show that the pattern persists in rolling windows ([Online Appendix Figure OA.2](#)), under compositionally adjusted wages (Panel B of [Online Appendix Table OA.2](#)), under alternative skill binning (Panel C), and under explicit top-tail censoring/winsorization choices (Panels D–E). Together, these checks reduce the concern that the headline fact is an artifact of changes in MSA coverage, upper-tail wage coding, or a particular hourly-wage construction

choice.⁹

2.3 Skill Premium by Skill Ratio over Time

The skill-premium fact in Section 2.3 shows that while the skill premium used to be lower in skill-abundant places, it has become higher in skill-abundant places in recent years. In Figure 3, I define the skill premium as the difference between the log wages of highly skilled and less skilled workers. I run the following regression:

$$\ln \left(\frac{\hat{w}_{Hjt}}{\hat{w}_{Ljt}} \right) = \sum_s \delta_s 1(t = s) (H_{jt}/L_{jt}) + \phi_j + \phi_t + \epsilon_{jt} \quad (2)$$

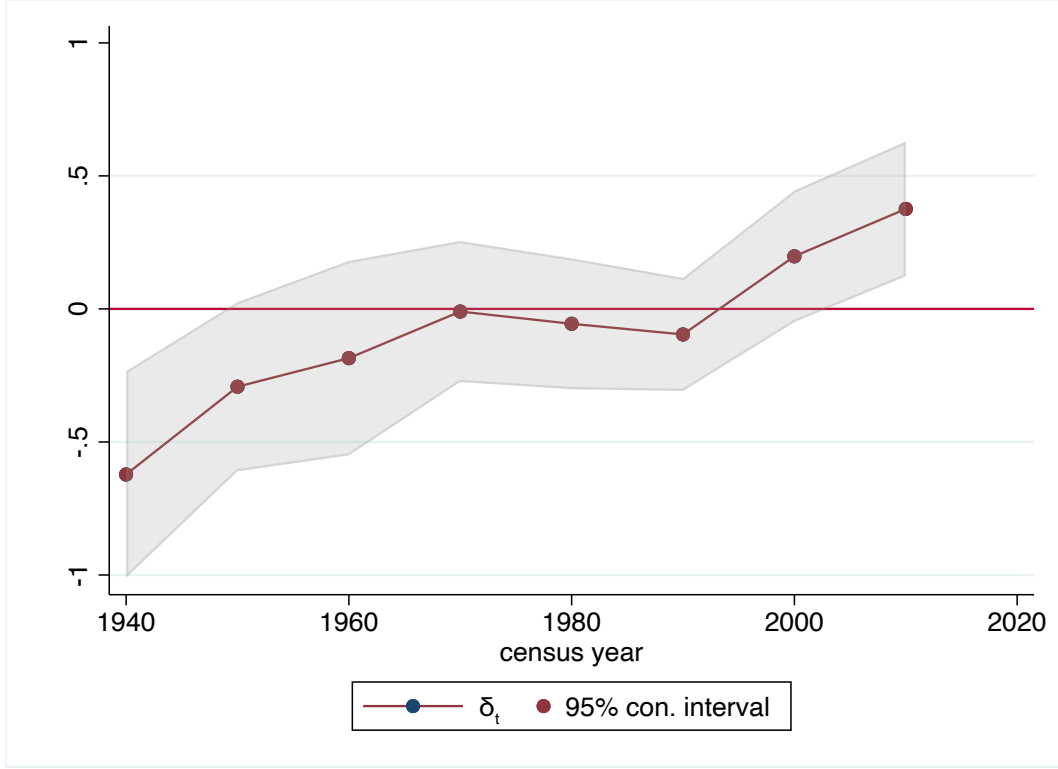
where $\hat{w}_{Hjt}$ and $\hat{w}_{Ljt}$ are the highly skilled and less skilled wages, respectively, constructed from the IPUMS data for MSA j at time t . ϕ_j is the MSA fixed effect, and ϕ_t is the year fixed effect. $\frac{H_{jt}}{L_{jt}}$ is the ratio of highly skilled to less skilled workers in MSA j at time t . Before estimation, I standardize within each year both the skill premium and the MSA skill-ratio regressor, so the reported δ_s are standardized slope coefficients. I estimate this pooled specification over the decennial observations from 1940 through 2010 by population-weighted OLS with MSA and year fixed effects and standard errors clustered at the MSA level. The indicator $1(t = s)$ selects the year-specific slope. This coefficient can be interpreted as follows: a one-standard-deviation increase in $\frac{H_{jt}}{L_{jt}}$ in year s changes the skill premium by δ_s standard deviations. Column (2) of Appendix Table A.3 adds raw population, and Figure 3 plots the column-(2) estimates. Baum-Snow and Pavan (2013) find that at least 23% of the overall increase in the variance of log hourly wages in the U.S. 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.

2.4 The Concentration of Highly Skilled Migrants

The third fact documents how exact one-year inter-MSA mover status varies with worker skill and the education composition of the worker's current MSA. I estimate a pooled probit with survey-year-specific main and interaction effects, as in equation (3).

⁹I identified this fact in the data in the first version of this paper in September 2016. In related work, Autor (2019) notes a decline in the urban less skilled wage premium. He does not report that regional wage divergence has occurred for the wages of highly skilled but not less skilled workers.

Figure 3: Skill Premium by MSA Education Levels



Note: Figure 3 plots the estimate of the coefficient δ_s for the specification (2). On the horizontal axis, I plot the decades from 1940 to 2010. On the vertical axis, I plot the estimates of the coefficient δ_s for each decade from 1940 to 2010. A horizontal line at zero is shown for reference. Coefficients are reported in Appendix Table A.3.

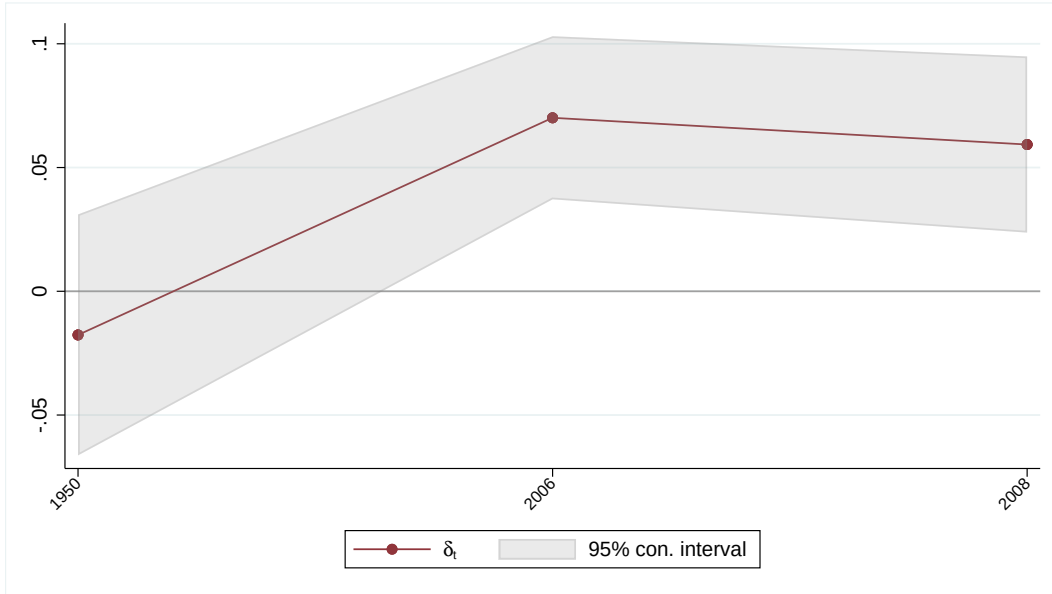
$$\begin{aligned} \Pr(\text{Migrant}_{ijt} = 1) = \Phi \left(\alpha + \sum_s 1(t = s) \left[\beta_s 1(H_{ijt}) + \gamma_s \left(\frac{H_{jt}}{L_{jt}} \right) \right. \right. \\ \left. \left. + \delta_s 1(H_{ijt}) \times \left(\frac{H_{jt}}{L_{jt}} \right) \right] + \Gamma X_{ijt} + \mu_{\text{state}(j)} + \lambda_t \right) \end{aligned} \quad (3)$$

where Migrant_{ijt} equals one when worker i 's current general MSA j differs from the comparable prior-year general MSA code. Respondents who report living in the same house are coded zero, while unidentified, foreign, and invalid prior locations are excluded. The term $1(t = s)$ selects survey year s , and $1(H_{ijt})$ equals one if the worker is highly skilled. $\frac{H_{jt}}{L_{jt}}$ is the contemporaneous skill ratio of the current MSA. Thus δ_s is a survey-year-specific association between inter-MSA mover status, worker skill, and current-MSA skill density; it is not a bilateral destination-choice coefficient. The vector X_{ijt} includes sex, age group, and the person-weight covariate. The fullest specification also includes state and survey-year fixed effects, and standard errors

are clustered by current MSA.¹⁰ I estimate (3) on employed individuals aged 25–55, pooling the exact one-year observations for 1950, 2006, and 2008. I restrict the sample to these three years because they are the years in the matched IPUMS files used in the analysis for which current MSA, one-year-prior MSA, and the required covariates are jointly available. I do not combine them with observations reporting residence five years earlier: one- and five-year mover indicators measure different cumulative migration horizons, and rescaling a five-year indicator does not recover a one-year migration outcome.

Figure 4 shows that the estimated interaction is small and statistically indistinguishable from zero in 1950, but positive in the 2006 and 2008 ACS samples. I use this exact one-year specification in the main text because it maintains the same migration horizon and geographic definition across survey years. Appendix Table A.4 reports all four specifications and the associated survey-year-specific main effects for this preferred outcome. Appendix Figure A.1 and Table A.5 report a complementary March CPS specification with broader historical coverage. Because the available CPS records do not maintain the same retrospective migration horizon in every period, I report that evidence separately rather than combine it with the one-year series.

Figure 4: One-Year Inter-MSA Migration by Destination Education Level



Note: Figure 4 plots the estimated survey-year-specific interaction coefficient δ_s from the fullest specification in equation (3) for exact one-year inter-MSA migration in 1950, 2006, and 2008. Shaded regions are 95% confidence intervals, and a horizontal line at zero is shown for reference.

¹⁰Appendix Section A.1, under “Migration Rates and Migrants,” gives the outcome and sample details.

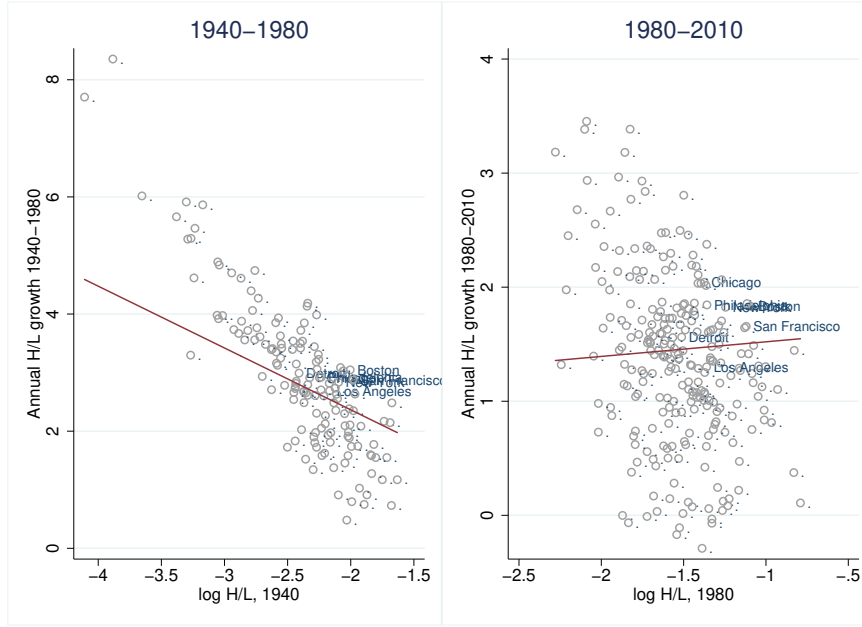
2.5 “The Great Convergence” and “The Great Divergence” before and after 1980

The final fact documented in this paper is that the skill ratio converged across MSAs before 1980, a pattern that mirrors the post-1980 “Great Divergence” of skills. [Moretti \(2004\)](#), [Berry and Glaeser \(2005\)](#), [Moretti \(2012\)](#), and [Diamond \(2016\)](#) show that skill ratios diverged across MSAs between 1980 and 2010. [Moretti \(2012\)](#) coins the term “The Great Divergence” to stress how the skills diverge over space. Nevertheless, what happened to the skill ratio before 1980? Was the skill distribution converging across MSAs when wages were converging? To answer this question, I analyze the convergence rates of the skill ratio over the last 70 years. I estimate the following specification:

$$\left(\ln \left(\frac{H_{jt}}{L_{jt}} \middle/ \frac{H_{j\tau}}{L_{j\tau}} \right) \right) \frac{1}{(t - \tau)} = \alpha + \beta^{HL} \ln \frac{H_{j\tau}}{L_{j\tau}} + \epsilon_{jt} \quad (4)$$

where, for $q \in \{\tau, t\}$, H_{jq} and L_{jq} denote the numbers of highly skilled and less skilled workers, respectively, living in MSA j in period q . The dependent variable is the average annual growth of the skill ratio between τ and t . With this regression, I can assess how growth in the skill ratio is related to the initial skill ratio. This regression is analogous to the regressions run in Figures 1 and 2 but for quantities rather than for wages. I run this regression over different periods using the Census and ACS data. In Figure 5, I plot the observations at the MSA level and then the fitted line, where β^{HL} convergence rates are the slopes of the lines. Each circle is an observation by MSA. I label the ten largest U.S. MSAs. Between 1940 and 1980, the β^{HL} -convergence rate was negative and statistically significant, motivating the term “Great Convergence”. However, as observed in [Moretti \(2004\)](#), between 1980 and 2010, the β^{HL} -convergence rate was positive and statistically significant, indicating divergence. Appendix Table A.6 reports estimates for shorter subperiods. Convergence was strongest in the early decades, followed by divergence from the 1970s through the 1990s and renewed convergence in 2000–2010.

Figure 5: Skill Convergence across MSAs before and after 1980



Note: Figure 5 plots each MSA's annual average skill growth (demeaned) against its (demeaned) initial skill level. The left panel depicts 1940–1980, and the right panel depicts 1980–2010. The red line depicts a weighted least-squares bivariate regression, with each MSA weighted by its population in the initial year.

3 A Spatial Equilibrium Model with Skills and Technology Diffusion

The empirical analysis indicates that long-run changes pushed the concentration of skills and their returns toward already highly skilled locations. Specifically, these changes are associated with patterns of regional convergence, the skill premium, and cross-MSA migration. These observations, along with the timing of such changes, suggest that a national skill-biased technology shock interacting with local agglomeration forces might rationalize the aforementioned long-term changes. To account for these effects, I develop a dynamic model of cities nesting the quantitative spatial equilibrium literature with spatial economic growth. To quantify the effect of technological progress and compare it with other forces, I also embed migration costs, costly trade across cities and with the Rest of the World, a housing sector, amenities, and industries into the model. The discussion at the end of this section explains the role of each additional feature in turn. Appendix Table B.1 summarizes the notation used in the structural model. Section 6 summarizes the principal robustness exercises, while Appendix Section D develops the complementary explanations in detail.

Environment. Time is discrete and indexed by decades, $t \in \{0, 1, \dots, T\}$. There is a finite set of domestic locations indexed by $j \in \{1, \dots, N\}$. There are two worker types, $k \in \{H, L\}$: highly skilled (H) and less skilled (L).

Intratemporal preferences and budget. In period t , the consumption component of a type- k worker's normalized preference index in location j is Stone–Geary, with tradable consumption T_{kjt} , housing O_{kjt} , and subsistence housing $\bar{O} > 0$.¹¹

$$u_{kjt}^C = \theta \ln(T_{kjt}) + (1 - \theta) \ln(O_{kjt} - \bar{O}), \quad \theta \in (0, 1). \quad (5)$$

The deterministic preference index combines consumption and location amenities:

$$\begin{aligned} \nu_{kjt}(T, O) &= u_{kjt}^C(T, O) + A_{kjt} + \gamma^p \ln(H_{jt}/L_{jt}) - c_\theta, \\ c_\theta &= \theta \ln \theta + (1 - \theta) \ln(1 - \theta). \end{aligned}$$

The common constant c_θ fixes the origin of the index. Here, A_{kjt} denotes amenities for skill group k in location j at date t , and $\ln(H_{jt}/L_{jt})$ is the natural logarithm of the ratio of highly skilled to less skilled workers in that location. The coefficient γ^p governs how local skill composition contributes to amenities. The within-period budget constraint for a type- k individual living in location j at time t is

$$P_{jt}T_{kjt} + R_{jt}O_{kjt} = W_{kjt},$$

where P_{jt} is the local price index of the tradable composite, R_{jt} is the rental rate of housing, and W_{kjt} is the nominal wage income of skill type k in location j at time t . For $W_{kjt} > R_{jt}\bar{O}$, maximizing this index subject to the budget constraint yields the normalized indirect preference index

$$\delta_{kjt} \equiv \log(W_{kjt} - R_{jt}\bar{O}) - \theta \log P_{jt} - (1 - \theta) \log R_{jt} + A_{kjt} + \gamma^p \ln\left(\frac{H_{jt}}{L_{jt}}\right), \quad (6)$$

The index δ_{kjt} is expressed in normalized location-choice units. For each skill group fix $a_k \in (0, 1)$, and let ζ_{ikjt} be independent across workers, locations and dates with the standard type-I extreme-value distribution $\Pr(\zeta \leq z) = \exp[-\exp(-z)]$. Define $\varepsilon_{ikjt} = \exp(a_k \zeta_{ikjt})$.

¹¹The preferences exhibit a degree of nonhomotheticity similar to [Ganong and Shoag \(2017\)](#). This allows the quantitative model to capture the interaction between wages, rents, and housing expenditures. The main qualitative predictions would continue to hold under homothetic preferences, but the Stone–Geary formulation provides a closer mapping to the housing mechanism in the data.

Period utility is $\exp\{a_k \nu_{kjt}(T, O)\} \varepsilon_{ikjt}$, so after consumption optimization it is

$$u_{ikjt} = \exp(a_k \delta_{kjt}) \varepsilon_{ikjt}. \quad (7)$$

This increasing transformation of the consumption index yields the Stone–Geary demands. The systematic component of log utility is $a_k \delta_{kjt}$.

Lifetime utility. Let $m_k(j, j') > 0$ denote the migration wedge in normalized choice units, with counterpart $M_k(j, j') = m_k(j, j')^{a_k}$ in utility levels. Following [Desmet, Nagy and Rossi-Hansberg \(2018\)](#), a move from j to j' permanently scales subsequent flow utility by $M_k(j, j')^{-1}$. For the migration history $\bar{j}_t = (j_0, \dots, j_t)$, flow utility is

$$\tilde{u}_{ikjt}(\bar{j}_t) = \left[\prod_{s=1}^t m_k(j_{s-1}, j_s)^{-a_k} \right] \exp(a_k \delta_{kjt}) \varepsilon_{ikjt}.$$

Initial locations are distributed according to the date-zero static equilibrium. Conditional on the given initial location j_0 , the worker chooses locations at dates $t = 1, \dots, T$ to maximize

$$U_{ik} = \mathbb{E} \left[\sum_{t=0}^T \beta^t \tilde{u}_{ikjt}(\bar{j}_t) \middle| j_0 \right], \quad \beta \in (0, 1).$$

The empty product is one, and location decisions use current and past taste shocks. At each decision date, earlier utility and inherited positive multipliers can be factored out of the remaining problem.

Assumption 1. *The normalized migration wedges are time-invariant and separable:*

$$m_k(j, j') = m_{k1}(j) m_{k2}(j'), \quad m_{k1}(j) > 0, \quad m_{k2}(j') \geq 1, \quad m_k(j, j) = 1. \quad (8)$$

Thus $m_{k1}(j) m_{k2}(j) = 1$ for every location.

Let $V_{kt}(j)$ be the remaining value in utility levels for a type- k worker entering decision date t in origin j , before drawing the current taste shocks and after removing inherited multipliers. Conditional on the aggregate path, with terminal value $V_{k,T+1}(j) = 0$, the Bellman equation is

$$V_{kt}(j) = \mathbb{E}_\zeta \left[\max_{j'} \left\{ \frac{\exp(a_k \delta_{kj't}) \varepsilon_{ikj't} + \beta V_{k,t+1}(j')}{m_k(j, j')^{a_k}} \right\} \right]. \quad (9)$$

Separability gives

$$V_{kt}(j) = \frac{1}{m_{k1}(j)^{a_k}} \mathbb{E}_\zeta \left[\max_{j'} \left\{ \frac{\exp(a_k \delta_{kj't}) \varepsilon_{ikj't} + \beta V_{k,t+1}(j')}{m_{k2}(j')^{a_k}} \right\} \right]. \quad (10)$$

Backward induction establishes

$$V_{kt}(j) = \frac{\Phi_{kt}}{m_{k1}(j)^{a_k}}, \quad \Phi_{k,T+1} = 0, \quad (11)$$

where Φ_{kt} is independent of the origin. Indeed, the continuation term in (10) satisfies

$$\frac{V_{k,t+1}(j')}{m_{k2}(j')^{a_k}} = \frac{\Phi_{k,t+1}}{[m_{k1}(j')m_{k2}(j')]^{a_k}} = \Phi_{k,t+1}.$$

Writing $B_{kt} = \sum_{s=1}^N \exp(\delta_{kst})/m_{k2}(s)$, the recursion is

$$\Phi_{kt} = \Gamma(1 - a_k)B_{kt}^{a_k} + \beta\Phi_{k,t+1},$$

where Γ denotes the gamma function. The maximum of $\delta_{kst} - \log m_{k2}(s) + \zeta_{ikst}$ has distribution $\log B_{kt} + G$, with G standard type-I extreme value and $\mathbb{E} \exp(a_k G) = \Gamma(1 - a_k) < \infty$. Thus expected utility is finite on a finite path with finite indices and positive wedges.

Since the continuation term is common across destinations, location choice solves

$$j_t \in \arg \max_j \left\{ \frac{\exp(a_k \delta_{kjt}) \varepsilon_{ikjt}}{m_{k2}(j)^{a_k}} \right\}. \quad (12)$$

Taking logs and dividing by $a_k > 0$ gives

$$j_t \in \arg \max_j \{ \delta_{kjt} - \log m_{k2}(j) + \zeta_{ikjt} \}. \quad (13)$$

Logit choice probabilities. The independent standard type-I extreme-value shocks imply

$$\pi_{kjt} \equiv \Pr(j_t = j \mid k, t) = \frac{\exp\{\delta_{kjt} - \log m_{k2}(j)\}}{\sum_{s=1}^N \exp\{\delta_{kst} - \log m_{k2}(s)\}}. \quad (14)$$

The fixed scale a_k determines utility levels and cancels from the location probabilities. In terms of the normalized index and destination components, the share system is

$$\pi_{kjt} = \frac{\exp(\delta_{kjt})/m_{k2}(j)}{\sum_{s=1}^N \exp(\delta_{kst})/m_{k2}(s)}. \quad (15)$$

Let H_t^{tot} and L_t^{tot} denote the aggregate supplies of highly skilled and less skilled workers at date t , respectively. Their allocations across locations are:

$$H_{jt} = H_t^{tot} \pi_{Hjt}, \quad L_{jt} = L_t^{tot} \pi_{Ljt}. \quad (16)$$

Equations (15)–(16) give location probabilities and populations at each decision date under time-invariant separable migration wedges. Amenities and the local skill-composition term determine time-varying attractiveness through the normalized index.

Technology. In this subsection, I describe the production technologies for the tradable composite good T , the set of sectoral intermediates $d \in \{1, \dots, D\}$, and the housing sector O . The final composite T_{jt} is assembled from traded sectoral aggregates $\{Q_{djt}\}_{d=1}^D$; each Q_{djt} is in turn a CES aggregator of location–sector varieties that use both highly skilled and less skilled labor; and housing O_{jt} is supplied locally as a function of rents. Throughout, productivity terms are allowed to depend on (i) agglomeration in the quantity and composition of local labor, (ii) skill-biased technical change, and (iii) an exogenous technology component that diffuses across space over time.

Final-good production. In each location j and period t , the tradable composite T_{jt} is a Cobb–Douglas aggregate of sectoral composites $\{Q_{djt}\}_{d=1}^D$:

$$T_{jt} = \prod_{d=1}^D \left(\frac{Q_{djt}}{\alpha_d} \right)^{\alpha_d}, \quad \sum_{d=1}^D \alpha_d = 1,$$

where α_d denotes the expenditure share on sector d . I normalize the final-good aggregator so that its price index is $P_{jt} = \prod_{d=1}^D P_{djt}^{\alpha_d}$. In the quantitative implementation, I assign equal expenditure shares across the D sectors, $\alpha_d = 1/D$.

Sector d in location j combines varieties imported from all origins i in a CES fashion:

$$Q_{djt} = \left[\sum_{i=1}^N a_{di} \left(y_{dijt} \right)^{\frac{\sigma_d - 1}{\sigma_d}} \right]^{\frac{\sigma_d}{\sigma_d - 1}}, \quad \sigma_d > 1,$$

where y_{dijt} is the quantity of the d -variety shipped from i to j , $a_{di} > 0$ are sector-origin-specific taste parameters, and σ_d is the elasticity of substitution across varieties. As long as $\sigma_d > 1$, consumers in j demand varieties from all origins; trade frictions and relative unit costs then shape the equilibrium sourcing pattern.

Intermediates, skills, and agglomeration forces. Each location–sector produces a continuum of varieties with the same technology and unit cost. Production combines highly skilled and less skilled labor through a CES technology:

$$Y_{djt} = \left[\eta_{Ldjt} L_{djt}^\rho + \eta_{Hdjt} H_{djt}^\rho \right]^{\frac{1}{\rho}}, \quad \forall j \in \{1, \dots, N\}, \quad d \in \{1, \dots, D\}, \quad t, \quad (17)$$

where H_{djt} and L_{djt} are, respectively, the quantities of high- and low-skill labor employed in sector d at location j and time t , and $\rho < 1$. The elasticity of substitution between H and L is $\sigma^{HL} \equiv 1/(1 - \rho) > 1$.

The skill-specific productivity shifters η_{Hdjt} and η_{Ldjt} capture agglomeration effects and exogenous technology. The literature on the social returns to education has documented that areas with a higher concentration of college graduates tend to exhibit higher wages and productivity. One plausible mechanism behind this pattern is knowledge spillovers, as explored in [Davis and Dingel \(2019\)](#), although alternative explanations such as sorting and network effects are also consistent with the observed correlations. Departing from a standard CES specification as in [Katz and Murphy \(1992\)](#), I allow these shifters to depend on (i) the local skill ratio H_{jt}/L_{jt} , (ii) total local population $H_{jt} + L_{jt}$, (iii) a skill-biased technical change component S_{kjt} , and (iv) baseline log productivity ξ_{kdjt} .¹²

$$\begin{aligned}\eta_{Hdjt} &= \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H} S_{Hjt}^{\Lambda_{Ht}} \exp(\xi_{Hdjt}), \\ \eta_{Ldjt} &= \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (H_{jt} + L_{jt})^{\phi^L} S_{Ljt}^{\Lambda_{Lt}} \exp(\xi_{Ldjt}),\end{aligned}\tag{18}$$

The parameters γ^H, γ^L capture productivity spillovers from the local skill ratio, while ϕ^H, ϕ^L capture spillovers from total city size. Consistent with the evidence in [Section 2](#), these elasticities allow the wages of highly skilled workers to respond more strongly to increases in both the skill ratio and city size than the wages of less skilled workers, matching the empirical patterns that (i) skill premia are increasingly higher in skill-abundant places and (ii) highly skilled workers are increasingly concentrated in skill-dense cities. Together, they capture the hypothesis that agglomeration forces are stronger in cities with higher concentrations of highly skilled workers. Simultaneously, following [Davis and Dingel \(2020\)](#) and [Baum-Snow, Freedman and Pavan \(2018\)](#), the spillover effects also appear for population, not just the skill ratio.¹³

The terms S_{Hjt} and S_{Ljt} are skill-biased technology shifters. The original exposure mapping and its time-varying skill-specific scale Λ_{kt} are held fixed throughout the counterfactuals. Log productivity ξ_{kdjt} evolves according to the diffusion-plus-innovation law below, while $\exp(\xi_{kdjt})$ is the corresponding positive productivity level.

¹²In the latest version of [Diamond \(2016\)](#), spillovers are not modeled with a parametric formulation, but more importance is given to utility spillovers. My analysis, however, benefits from modeling productivity spillovers with specific functional forms, especially for the counterfactual analysis.

¹³[Section 3.2](#) states an existence result for the baseline static equations and a separate conditional uniqueness criterion.

Technology diffusion and realized innovations. Following [Desmet, Nagy and Rossi-Hansberg \(2018\)](#), persistent log productivity combines its own lag with a common diffusion frontier. The corresponding law of motion is:

$$\xi_{k d j t} = \varrho_k \xi_{k d j, t-1} + (1 - \varrho_k) \log \left(\frac{1}{N} \sum_{s=1}^N \exp(\xi_{k d s, t-1}) \right) + \varepsilon_{k d j t}, \quad k \in \{H, L\}. \quad (19)$$

Here, $\xi_{k d j, t-1}$ is location j 's lagged log productivity, and $\log \left(\frac{1}{N} \sum_{s=1}^N \exp(\xi_{k d s, t-1}) \right)$ is the log of lagged average productivity in levels across locations for skill k and sector d . The coefficients ϱ_k and $1 - \varrho_k$ govern the balance between local persistence and diffusion: when ϱ_k is higher, productivity depends more strongly on its own local history; when it is lower, productivity is pulled more strongly toward the common component. The term $\varepsilon_{k d j t}$ is a location-, sector-, skill-, and date-specific innovation in log productivity; equivalently, $\exp(\varepsilon_{k d j t})$ shifts the positive productivity level multiplicatively.

Firm behavior and wages. Each variety producer in sector d at location j and time t takes prices and productivity as given and chooses $(H_{d j t}, L_{d j t})$ to maximize static profits,

$$\pi_{d j t} = \max_{h, l} \left\{ p_{d j t} \left[\eta_{L d j t} l^\rho + \eta_{H d j t} h^\rho \right]^{\frac{1}{\rho}} - W_{H j t} h - W_{L j t} l \right\},$$

where $p_{d j t}$ is the factory–gate price and $(W_{H j t}, W_{L j t})$ are wages for high– and low–skill labor. Under perfect competition and free entry, profits are driven to zero, so firms are indifferent over scale and only unit costs matter. The first–order conditions equate wages to marginal products of labor, yielding

$$W_{H j t} = p_{d j t} \eta_{H d j t} \left[\eta_{L d j t} L_{d j t}^\rho + \eta_{H d j t} H_{d j t}^\rho \right]^{\frac{1}{\rho}-1} H_{d j t}^{\rho-1}, \quad (20)$$

$$W_{L j t} = p_{d j t} \eta_{L d j t} \left[\eta_{L d j t} L_{d j t}^\rho + \eta_{H d j t} H_{d j t}^\rho \right]^{\frac{1}{\rho}-1} L_{d j t}^{\rho-1}. \quad (21)$$

Dual unit cost. Let $\sigma^{HL} \equiv \frac{1}{1-\rho} > 1$ denote the elasticity of substitution between H and L in (17). Given $(W_{H j t}, W_{L j t})$ and productivity $\eta_{k d j t}$, the unit cost for sector d in origin j is

$$c_{d j t}^{1-\sigma^{HL}} = \eta_{L d j t}^{\sigma^{HL}} W_{L j t}^{1-\sigma^{HL}} + \eta_{H d j t}^{\sigma^{HL}} W_{H j t}^{1-\sigma^{HL}}, \quad (22)$$

which is the only object that matters for equilibrium trade flows and price indices.

Costly interregional trade of intermediates. Intermediate goods in sector d are traded across locations with iceberg trade costs $\tau_{jj'}^d \geq 1$, where $\tau_{jj}^d = 1$ for all j . If one unit of the d -variety produced in j is shipped to j' , only $1/\tau_{jj'}^d$ units arrive. Let $p_{dj't}$ denote the sector-specific producer price in origin j . The trade system uses sectoral Armington price indices and sector-specific producer prices, solved industry by industry. For tractability, the quantitative implementation normalizes Armington origin weights to one within each sector. Sectoral price indices therefore satisfy

$$P_{dj't} = \left(\sum_{j=1}^N \left(\tau_{jj'}^d p_{dj't} \right)^{1-\sigma_d} \right)^{\frac{1}{1-\sigma_d}}, \quad (23)$$

and the corresponding bilateral expenditure shares are

$$\pi_{jj't}^d = \left(\tau_{jj'}^d p_{dj't} \right)^{1-\sigma_d} \left(P_{dj't} \right)^{\sigma_d-1}. \quad (24)$$

Sector- d varieties are traded across space, while the final tradable bundle $T_{j't}$ is assembled locally and is non-traded. Under Stone–Geary preferences, workers' tradable expenditure in location j' is

$$X_{j't}^W = \theta \left[(W_{Hj't} - R_{j't}\bar{O})H_{j't} + (W_{Lj't} - R_{j't}\bar{O})L_{j't} \right].$$

Housing is locally owned, and landlords spend rental income $R_{j't}^{\mu+1}$ on the local tradable bundle. Total domestic absorption is therefore $\mathcal{A}_{j't} = X_{j't}^W + R_{j't}^{\mu+1}$. Using the Cobb–Douglas expenditure shares of the final-good aggregator, implied domestic expenditure on sector d is $E_{dj't} = \alpha_d \mathcal{A}_{j't}$ and implied bilateral trade flows are $X_{jj't}^d = \pi_{jj't}^d \alpha_d \mathcal{A}_{j't}$. The composite price index for the tradable good $T_{j't}$ is then

$$P_{j't} = \exp \left(\sum_{d=1}^D \alpha_d \log P_{dj't} \right),$$

which is the object entering the indirect utility function (6). Producer prices equal unit costs, $p_{dj't} = c_{dj't}$, and sectoral revenue equals domestic demand plus demand from the Rest of the World (ROW):

$$p_{dj't} Y_{dj't} = \sum_{j'=1}^N \pi_{jj't}^d \alpha_d \mathcal{A}_{j't} + X_{j't}^{d,ROW}. \quad (25)$$

When the ROW is active, reciprocal import and export shares are computed with the same Armington system and ROW expenditure is chosen so that aggregate exports equal aggregate imports. Thus sectoral goods markets, factor markets, the domestic budget including local housing income, and the external trade balance all clear jointly.

Housing market. Housing is a non-tradable sector produced locally in each location j with a reduced-form technology that maps rental prices into quantities of housing services:

$$O_{jt} = R_{jt}^\mu, \quad (26)$$

where R_{jt} is the rental rate and $\mu > 0$ is the elasticity of housing supply.¹⁴ Aggregate housing demand in location j at time t is the sum of individual housing demands from highly skilled and less skilled workers, derived from the Stone–Geary preferences in (5). Individual housing demand for a type- k worker is

$$\theta \bar{O} + (1 - \theta) \frac{W_{kjt}}{R_{jt}}, \quad (27)$$

where $\bar{O}$ is the subsistence level of housing and θ is the marginal expenditure share on tradables out of supernumerary income. Housing-market clearing then requires local supply to equal the sum of those individual demands across residents:

$$R_{jt}^\mu = H_{jt} \left[\theta \bar{O} + (1 - \theta) \frac{W_{Hjt}}{R_{jt}} \right] + L_{jt} \left[\theta \bar{O} + (1 - \theta) \frac{W_{Ljt}}{R_{jt}} \right]. \quad (28)$$

For the baseline values $\mu \in (0, 1)$, the housing-supply schedule is increasing and concave in rents, so an increase in local income or an influx of migrants raises R_{jt} and dampens additional in-migration.

3.1 Equilibrium

At each date t , the state of the economy is given by log productivity $\{\xi_{kdt}\}_{k,d,j}$ together with the fundamentals $\{A_{kjt}, S_{kjt}, \tau_{jj'}^d, m_k(j, j')\}$. Given these objects, the preference, housing, production, and trade parameters, and the aggregate supplies of each skill (H_t^{tot}, L_t^{tot}) , a *static competitive spatial equilibrium* at date t is a vector of prices and allocations such that (i) households' location and within-location choices are optimal, (ii) firms minimize costs and price competitively, and (iii) labor, goods, and housing markets clear in every location. A *dynamic spatial equilibrium* is a sequence of such static equilibria over time, linked by the technology diffusion process (19). The formal definitions follow.

Definition 1 (Static competitive spatial equilibrium at date t). Fix a date t . Take as given log productivities $\{\xi_{kdt}\}_{k,d,j}$, amenities $\{A_{kjt}\}_{k,j}$, SBTC shifters $\{S_{kjt}\}_{k,j}$, iceberg trade costs

¹⁴A fully dynamic housing model with investment decisions along the lines of Glaeser et al. (2014) is a possible extension of the paper. However, to avoid moving the paper's focus away from skill-biased technology, I keep the housing market as simple as possible. This simplification also enhances comparability to Ganong and Shoag (2017). I run some simulations, varying μ over a wide range, including very low values, to see how housing would respond.

$\{\tau_{jj'}^d\}_{d,j,j'}$, migration-cost primitives $\{m_k(j, j')\}_{k,j,j'}$, the preference and housing parameters $(\beta, \theta, \bar{O}, \gamma^p, \mu)$, the production and trade parameters $(\rho, \{\sigma_d\}_d, \gamma^H, \gamma^L, \phi^H, \phi^L)$, and aggregate skill supplies (H_t^{tot}, L_t^{tot}) .

A static competitive spatial equilibrium at date t is a collection of prices and allocations $\{W_{kjt}, R_{jt}, P_{jt}, P_{djt}, H_{jt}, L_{jt}, H_{djt}, L_{djt}, X_{jj't}^d, Y_{djt}\}_{k,d,j,j'}$ such that:

1. **Households.** Given $(W_{kjt}, P_{jt}, R_{jt})$, indirect utility δ_{kjt} is given by (6). The population allocations (H_{jt}, L_{jt}) satisfy the location-choice system (16), and aggregate feasibility holds: $\sum_j H_{jt} = H_t^{tot}$, $\sum_j L_{jt} = L_t^{tot}$.
2. **Productivity.** Skill- and sector-specific productivity shifters satisfy (18) given (H_{jt}, L_{jt}) , S_{kjt} , and ξ_{kdjt} .
3. **Firms and labor markets.** Production satisfies (17). Unit costs satisfy (22), and perfect competition implies zero profits. Wages satisfy (20)–(21). Sectoral labor demands add up to total local employment:

$$H_{jt} = \sum_{d=1}^D H_{djt}, \quad L_{jt} = \sum_{d=1}^D L_{djt}. \quad (29)$$

4. **Trade and goods markets.** Sectoral price indices satisfy (23), implied bilateral expenditure shares satisfy (24), and producer-price updates follow the trade closure in (25).
5. **Housing.** Housing supply satisfies (26), individual housing demand satisfies (27), and housing markets clear in each location according to (28).

Definition 2 (Dynamic spatial equilibrium). Given initial log productivities $\{\xi_{kdj0}\}_{k,d,j}$, fixed realized innovations $\{\varepsilon_{kdjt}\}_{k,d,j,t}$, and sequences $\{A_{kjt}\}_{k,j,t}$, $\{S_{kjt}\}_{k,j,t}$, and $\{(H_t^{tot}, L_t^{tot})\}_{t=0}^T$, a dynamic spatial equilibrium is obtained by iterating over $t = 0, \dots, T$:

- (i) **Static step:** given $\{\xi_{kdjt}\}_{k,d,j}$ and contemporaneous fundamentals, allocations and prices constitute a static competitive spatial equilibrium in the sense of Definition 1.
- (ii) **State transition:** log productivities update according to the diffusion law (19) for all $k \in \{H, L\}$, d , and j (for $t \geq 1$).

Thus, the date- t technology state is predetermined when the date- t static equilibrium is solved; after that equilibrium is obtained, the diffusion law constructs the technology state used to solve the date- $(t + 1)$ equilibrium.

See Appendix Section B.1 for additional details.

3.2 Existence and conditional uniqueness of equilibrium

Under the conditions in Proposition 1 in Appendix B.2, the baseline model without physical capital admits a static equilibrium with both skills in every city and strictly positive supernumerary income. The sufficient restriction on agglomeration is $\gamma^H + \gamma^p \phi^H < 2 - \rho$. Existence holds at every sufficiently large aggregate labor-income level for the equations without an additional domestic price or income restriction. Appendix B.3 provides the proof.

The conditional uniqueness result and its scope for the quantitative model are set out in Appendix B.2, which also discusses the construction of the equilibrium path.

3.3 Other Model Features

The model combines three central components: (i) spatial technology diffusion, which generates a force toward regional convergence; (ii) SBTC, which shifts relative demand toward highly skilled workers; and (iii) local productivity and amenity agglomeration together with endogenous sorting, which transmit these technological changes unevenly across cities and can generate regional divergence. The model also includes housing, migration frictions, trade across cities and with the Rest of the World, and multiple industries. These features determine how local wages, prices, and worker location choices adjust, and they allow the analysis to assess other potential explanations for the decline in regional convergence. I describe their roles below, while Appendix Section D provides further details.

Housing Sector. A change in housing regulation in more expensive locations might reduce regional convergence, as pointed out in the pioneering paper by Ganong and Shoag (2017). Ganong and Shoag (2017) suggest that the U.S. states where housing prices increased the most are also where migration declined. Hence, because migration increases convergence, the decline in migration to these areas, which are also the richest, decreased the income convergence rate. They suggest that housing prices and SBTC can be complementary. To disentangle these mechanisms, I add a housing sector to the model to compare the housing effects with my fundamental mechanisms. I also conduct additional empirical tests in Appendix Section D.

Multiple Industries. I include multiple industries for two main reasons. The first reason is that the identification strategy I describe in the estimation section exploits the industry structure of MSAs. Thus, including industries in the model keeps the interpretation of the instrument within the model. The second reason is that the industrial composition across U.S. cities is very heterogeneous and highly skilled workers are allocated differently across sectors and locations. This could have directly affected regional convergence. In Appendix

Section D, I discuss this possibility in further detail, including capital-skill complementarity and related tests.

Migration Costs. A decline in migration costs may also change the spatial allocation of workers and the concentration of highly skilled labor. To consider this mechanism, the Free Migration counterfactual preserves the estimated destination components of the bilateral migration costs through 1980 and sets them to one at decennial model dates from 1990 through 2010. As reported below in Section 5, the results on β -convergence change significantly in a counterfactual with free migration.

Amenities. An essential feature of geography to explain the “Great Divergence” of skills and several other aspects of regional differences is amenities, as studied by Diamond (2016). Thus, I must include and estimate them, since they may also have affected regional convergence, prompting the reallocation of highly skilled labor to skill-abundant areas. Amenity-induced inflows of highly skilled workers put downward pressure on their wages through increased labor supply, but can also raise productivity through agglomeration.

4 Taking the model to the Data

This section summarizes the data sources, measurement choices, estimation, inversion, and validation exercises used to take the model to the data, with further details in Appendix Sections A and C.

Data for Quantification. To bring the model to the data, I employ the data described in Section 2 across U.S. MSAs on wages by skill, levels of less skilled labor L and highly skilled labor H , and rents in each location at ten-year intervals from 1960 to 2010. I recover the log-productivity panel ξ_{kajt} , equivalently the positive levels $\exp(\xi_{kajt})$, and amenity vectors A_{kjt} at the six observed dates, while the time-invariant destination components $m_{k2}(j)$ of the bilateral migration costs enter as fixed normalized inputs. In addition, I need estimates of the sectoral elasticity of substitution across varieties σ_d and trade costs $\tau_{jj'}^d$. I also need estimates of the expenditure share of housing in consumption $(1 - \theta)$ and the subsistence level of housing $\bar{O}$. I discipline the elasticities that govern productivity agglomeration and utility-amenity spillovers $(\gamma^L, \gamma^H, \phi^H, \phi^L, \gamma^p)$ and estimate the two skill-specific technology-persistence parameters (ϱ_H, ϱ_L) . To measure skill-biased technical change, I use data on labor by occupation and skill and routine-intensive activities by occupation.

The quantitative model is evaluated at decadal intervals from 1960 to 2010, with 1960 as the initial period.¹⁵ I invert the static model at every observed date to recover the positive productivity levels $\exp(\xi_{kdt})$ and hence ξ_{kdt} , then recover the realized innovations that reconcile adjacent log productivities with equation (19). Amenities are inverted once for each skill group at each of the six dates after the centered-wage reconstruction. The same six-date amenity values are used in every counterfactual. I set the sectoral elasticity of substitution across varieties to $\sigma_d = 4.50$ for all d , following [Caliendo, Dvorkin and Parro \(2019\)](#). In the gravity system, this implies a trade elasticity of $\sigma_d - 1 = 3.50$.

The static parameters governing local agglomeration forces and endogenous amenities are estimated using a generalized method of moments (GMM) procedure that exploits within-country variation in wages, employment, rents, and migration over time. Technology persistence is estimated separately on the inverted positive city–sector productivity panel $\exp(\xi_{kdt})$ using the same transition operator as the simulator. I choose internal estimation (rather than importing values from other contexts) because these parameters are central for the quantitative counterfactuals and are likely to be environment- and period-specific. Moreover, a structural identification strategy is required to distinguish the contribution of local propagation mechanisms from that of exogenous shifts in labor demand.

The key empirical object is ΔS_{kjt} , which captures changes in SBTC at the location level. To proxy for it, I use changes in routine occupational shares that differentially affect highly skilled and less skilled workers. This yields indices ΔS_{Hjt} and ΔS_{Ljt} .

These measures have two important features. First, they are unit-free: the difference between a one-unit change in ΔS_H and ΔS_L is not directly interpretable in log-productivity units. In the model, the skill-specific scaling term Λ_{kt} maps these indices into productivity units. The alternative exposure specifications use skill-specific endpoint coefficients, as described in the paragraph on quantitative SBTC scaling parameters. Second, ΔS_{kjt} is potentially endogenous to local equilibrium adjustments (e.g., occupational reallocation). I address this second issue by instrumenting ΔS_{kjt} with a Bartik-style predictor, denoted $\Delta \hat{Z}_{kjt}$, constructed from local industry composition and a leave-one-out measure of routine intensity within industries, as described below.

This separation between local exposure and SBTC strength is central for the quantification. The first-step GMM identifies the static agglomeration and amenity parameters, while productivity persistence is estimated separately. The quantitative exercises use the baseline SBTC exposure path and use four skill-specific endpoint coefficients to construct alternative exposure specifications. Aggregate wage-growth and skill-premium moments provide the fit

¹⁵Unlike the empirical data section 2, where the analysis starts in 1940, I use 1960 as the starting point to avoid war and immediate post-war disruptions and to work with broader data availability across model inputs. All the data-model comparisons are done starting in 1960.

comparisons reported in Appendix Table C.4.

Regions, sectors, and labor markets The quantitative model employs U.S. MSAs and a constructed Rest of the World region. I consider 14 sectors, classified according to the 1990 Census Bureau industrial classification as defined in Appendix Table C.1. I exclude the agriculture, mining, public-administration, and armed-forces sectors; utilities are included as a non-tradable sector. Sectors are then divided into tradable and non-tradable. Agents can freely switch across sectors; thus, wages equalize across them.

Skill-Biased Technical Change. I measure local exposure to SBTC using the routinization approach of Autor and Dorn (2013). Following their logic, I classify occupations by routine task intensity (RTI) and construct the skill-specific local exposure measures ΔS_{Ljt} and ΔS_{Hjt} from changes in the employment shares of routine occupations for less skilled and highly skilled workers.¹⁶ In these empirical exposure equations, the time subscript t denotes the calendar year rather than the model period index, so $t - 10$ refers to the previous decade.

$$\Delta S_{Ljt} = \sum_{v=1}^{\Upsilon} \left(\frac{L_{jvt}}{L_{jt}} - \frac{L_{jvt-10}}{L_{jt-10}} \right) 1(RTI_v > RTI_{P66}) \quad (30)$$

$$\Delta S_{Hjt} = \sum_{v=1}^{\Upsilon} \left(\frac{H_{jvt}}{H_{jt}} - \frac{H_{jvt-10}}{H_{jt-10}} \right) 1(RTI_v > RTI_{P66}) \quad (31)$$

These skill-specific routinization measures proxy how SBTC affects cities differently depending on their occupational composition.¹⁷ Appendix Figure C.1 shows their aggregate evolution over time.

Because ΔS_{Ljt} and ΔS_{Hjt} may be correlated with contemporaneous local equilibrium adjustments, I instrument them with a Bartik-style predictor that combines lagged local industry composition with a leave-one-out measure of routine intensity across cities within industries. For $k \in \{H, L\}$, I write this predictor as

$$\Delta \hat{Z}_{kjt} = \sum_d \omega_{jd,t-10} \bar{R}_{dt}^{k,-j},$$

¹⁶Appendix Section D.4 reports results with alternative definitions of SBTC, allowing SBTC to vary by location only, by industry only, or only at the national level.

¹⁷While this approach provides a good proxy for the local impact of SBTC, it might not be the only one. Computer adoption might represent the arrival of computers and demonstrate how cities are affected differently. Beaudry, Doms and Lewis (2010) use this approach. However, the available data end in 2000. This shortfall prevents me from recreating the full analysis through 2010 and is insufficient to estimate my model. For this reason, I use the empirical approach of Autor and Dorn (2013), which is flexible and data-driven. It allows me to build an index for all years in the analysis.

where d indexes industries and t denotes calendar time in years. The list of industries d used to construct the measures is reported in Table C.1 in the Appendix.¹⁸ Here, $\omega_{jd,t-10}$ is the lagged employment share of industry d in MSA j at time $t - 10$, and $\bar{R}_{dt}^{k,-j}$ is the leave-one-out routine share for skill group k in industry d and year t , computed excluding city j from the cross-city average. Relative to Autor and Dorn (2013), I construct both the index and the instrument separately for highly skilled and less skilled workers. In this way, I generate additional skill-specific variation and use the differential impact of technological shocks across the two groups in the estimation. These instrumental variables, $\Delta\hat{Z}_{Ljt}$ and $\Delta\hat{Z}_{Hjt}$, are useful in the estimation of the parameters of the model and the construction of the moment conditions.

Appendix Table C.2 reports the first-stage coefficients and standard errors for these instrumental variables. The predictive relationship between ΔS_H and $\Delta\hat{Z}_H$ is sizable and highly significant across decades (Panel A), implying strong first-stage relevance. For ΔS_L (Panel B), first-stage relevance is also sizable and statistically significant, but implied single-instrument F-statistics are weaker in some early decades: below 10 in 1950, 1970, and 1980, with 1970 below 7.¹⁹

This two-step treatment of SBTC is preferable to estimating the aggregate SBTC scaling parameters directly from the shift-share variation used in the first-step GMM. First, Appendix Table C.2 shows that the first-stage strength varies substantially across decades and falls below conventional weak-IV thresholds in some periods. Second, the recent shift-share econometrics literature emphasizes that Bartik designs are effectively identified off the underlying shock dimension (e.g., industries or occupations), so the effective variation can be low-dimensional and concentrated in a small subset of shocks, making regression-based estimation fragile and potentially weakly identified (Adão, Kolesár and Morales, 2019; Borusyak, Hull and Jaravel, 2022; Goldsmith-Pinkham, Sorkin and Swift, 2020). For this reason, I use the first-step GMM to recover local propagation parameters, specify the endpoint coefficients for alternative exposures separately, and assess aggregate fit using national moments.

SBTC slowdown after 1980. The mechanism does not require an acceleration of SBTC after 1980: local agglomeration forces can propagate earlier technological changes even as the aggregate pace of SBTC slows. This is consistent with the slowdown in SBTC after 1970 documented by Acemoglu and Autor (2011). To examine the role of current and past exposure in the data, I include contemporaneous and lagged SBTC exposure measures in the augmented convergence regression reported in Online Appendix Section OA.1.3,

¹⁸The same industries d are also used to calibrate the model for internal consistency.

¹⁹As a robustness test, I estimate the model without the 1950s, and the parameter estimates are unchanged.

equation (OA.1), and Online Appendix Table OA.3. Across the nested specifications, the estimated β coefficient becomes progressively more negative as SBTC controls are added. The richest specification remains statistically indistinguishable from zero at conventional levels, although several exposure coefficients—especially the 1970 less-skilled measure—are statistically significant. The exercise therefore provides suggestive rather than decisive evidence of persistent propagation through local agglomeration dynamics. Appendix E characterizes wage and skill-premium changes from a common initial benchmark under a joint change in national skill bias and local productivity shifters, holding populations fixed. The premium-change gradients are independent of the national skill-bias factor; the wage-change gradients have the same signs under the stated support conditions. The full quantitative model evaluates the resulting sorting and regional convergence dynamics. As an additional model-based robustness test of this mechanism, Table D.7 reports a counterfactual in which the model input S_{Hjt} is set 15% above its observed value. The model then implies that β^H would have moved further from the baseline under a stronger SBTC process.

Exogenous technology inputs versus endogenous propagation. The quantitative exercise takes as given a set of empirically measured SBTC exposure series $\{\Delta S_{kjt}\}$ that capture the heterogeneous incidence of national technological change across local labor markets. The model translates these observed shifters into equilibrium wage dynamics through (i) endogenous migration, (ii) local productivity and utility agglomeration, and (iii) the persistent productivity transition. Aggregate fit is assessed using cumulative wage growth for both skill groups and the evolution of the skill premium, as reported in Appendix Table C.4. The productivity residual b_{kjt} then reconstructs the baseline spatial wage geography. Thus, the baseline β^H and β^L paths are implied by the inversion, rather than independent tests of the model. The mechanism exercise instead changes one structural mechanism at a time while holding the realized productivity innovations, productivity residuals, and amenity path fixed. Appendix D.4 reports the alternative SBTC measures, and Appendix D.8 shows that eliminating diffusion is an extreme whole-horizon sensitivity rather than a close robustness case.

Agglomeration, Endogenous Amenities and Productivity Diffusion. I estimate the static labor-demand and labor-supply parameters by GMM, following the logic of Diamond (2016), and estimate productivity persistence in a separate second step on the inverted positive city–sector panel. The next two subsections report the GMM identification strategies for $\{\gamma^L, \gamma^H, \rho, \phi^H, \phi^L\}$ and $\{\theta, \gamma^p\}$. The persistence design for $\{\varrho_H, \varrho_L\}$ is described after the static GMM results. Keeping these stages separate avoids imposing a diffusion law on the

city-level composite residual used by the static wage equations.

Labor Demand. Because wages are equalized across industries within a city and the propagation coefficients are common across sectors, the sectoral first-order conditions can be aggregated exactly to city-level wage equations; Appendix C.2 provides the derivation. Under that aggregation, sectoral prices and sectoral outputs do not enter the empirical wage equations separately. Instead, they are absorbed into a city-level composite residual, denoted Ξ_{kjt} , that summarizes the contribution of sectoral prices, sectoral activity, and sector-specific productivity components to the city-level wage of skill group k . Separate sectoral price data are therefore not required for the GMM step. The resulting city-level estimating equations are:

$$\Delta w_{Ljt} = (\rho - 1)\Delta \ln L_{jt} + \gamma^L \Delta \ln \frac{H_{jt}}{L_{jt}} + \phi^L \Delta \ln (H_{jt} + L_{jt}) + \zeta_L \Delta \ln S_{Ljt} + \Delta \Xi_{Ljt}$$

$$\Delta w_{Hjt} = (\rho - 1)\Delta \ln H_{jt} + \gamma^H \Delta \ln \frac{H_{jt}}{L_{jt}} + \phi^H \Delta \ln (H_{jt} + L_{jt}) + \zeta_H \Delta \ln S_{Hjt} + \Delta \Xi_{Hjt}.$$

The city-level composite changes $\Delta \Xi_{kjt}$ enter the static estimating equations as residual demand shifts. I denote the resulting city-level GMM disturbances by Δu_{kjt} and impose the labor-demand moment conditions on those disturbances; I do not impose the sectoral diffusion law on Ξ_{kjt} . The coefficients ζ_H and ζ_L are nuisance coefficients in these first-step estimating equations: they absorb the direct reduced-form association between the measured SBTC proxies and local wage changes. They are distinct from both the city-skill-date productivity residual b_{kjt} and the quantitative scaling term Λ_{kt} . After the GMM, persistence is estimated directly on the positive city-sector productivity panel, where the model transition is defined. In the spirit of [Diamond \(2016\)](#) and [Suárez Serrato and Zidar \(2016\)](#), the static identification strategy uses labor-supply changes that are uncorrelated with local productivity disturbances. The interaction of SBTC shocks with the cities' housing supply elasticities generates differential labor-supply responses to a common labor-demand shock. For example, consider two cities: one has very elastic housing supply and the other very inelastic housing supply. If both experience an increase in labor demand, workers move to take advantage of these increases. Once they move, however, the MSA with more inelastic housing has a larger rent increase, which dampens in-migration. The resulting smaller labor-supply response limits the equilibrium wage offset that would otherwise occur as workers move in.²⁰ Specifically, the

²⁰I supplement the SBTC exposure proxy with a classic Bartik-style instrument.

exclusion restrictions are:

$$E(\Delta u_{Hjt} \Delta \mathcal{Z}_{jt}) = 0 \quad \text{and} \quad E(\Delta u_{Ljt} \Delta \mathcal{Z}_{jt}) = 0$$

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

where x^{reg} and x^{unav} denote local land-regulation restrictions and land unavailability, respectively. These moment conditions identify the labor-demand parameters by using exogenous variation in local labor supply generated by differential housing-supply elasticities, and they are jointly estimated using a two-step GMM procedure. This is the point at which the housing block matters for identification: interacting exogenous labor-demand shocks with heterogeneous housing-supply elasticities generates variation in local employment and wages that is orthogonal to unobserved productivity shocks, following [Diamond \(2016\)](#) and [Suárez Serrato and Zidar \(2016\)](#).

Labor Supply. The labor-supply estimation uses the normalized indirect preference index δ_{kjt} in equation (6). Under time-invariant separable migration wedges, location probabilities take the static form derived above. Logit inversion recovers $\delta_{kjt} - \log m_{k2}(j)$ up to a skill–date normalization, following [Berry, Levinsohn and Pakes \(2004\)](#). First differencing removes the time-invariant destination component and gives:

$$\begin{aligned} \Delta \delta_{kjt} = & \ln \left(\frac{W_{kjt} - R_{jt} \bar{O}}{W_{kjt-10} - R_{jt-10} \bar{O}} \right) - \theta \ln \left(\frac{P_{jt}}{P_{jt-10}} \right) \\ & - (1 - \theta) \ln \left(\frac{R_{jt}}{R_{jt-10}} \right) + \gamma^p \ln \left(\frac{H_{jt}/L_{jt}}{H_{jt-10}/L_{jt-10}} \right) + \Delta A_{kjt} \end{aligned} \quad (32)$$

The parameters θ and γ^p are identified from normalized preference-index responses to wages, rents and the local skill ratio, and do not depend on $m_{k2}(j)$. The heterogeneous destination components of bilateral migration costs, $m_{k2}(j)$, are accounted for when inverting the location-choice system to recover amenities from population data, as in the later subsection. Identifying workers' preferences for wages, rent, non-tradable local goods, housing, and amenities requires variation in these MSA characteristics that is uncorrelated with local unobservable amenities ΔA_{kjt} . This reasoning follows [Diamond \(2016\)](#). Specifically, I use SBTC shocks and their interaction with the characteristics of the supply elasticity. For the exclusion restriction to be satisfied, the set of instruments needs to be uncorrelated with unobserved exogenous changes in the MSAs' local amenities. The key idea is that since national changes in industrial productivity drive SBTC shocks, these shocks are unrelated to changes in local exogenous amenities, as well as to the same shocks interacted with local land regulation x^{reg} and land

unavailability x^{unav} . Specifically, the moment restrictions are:

$$E(\Delta A_{Hjt} \Delta \mathcal{Z}_{jt}) = 0 \quad \text{and} \quad E(\Delta A_{Ljt} \Delta \mathcal{Z}_{jt}) = 0$$

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

I estimate the static parameters by GMM and report the results in Table C.3. Panels A and B report the structural propagation parameters. The nuisance coefficients (ζ_H, ζ_L) on the SBTC proxies are included in the wage-change equations but omitted from the table because they are not structural objects used in the quantitative simulations. Panel C reports the separately estimated persistence parameters.

Estimation Results and Robustness: Agglomeration Forces I use a GMM estimation procedure with MSA-level data for 1940–2010, with estimates produced every ten years. The estimation sample starts earlier than the simulation window; the dynamic model itself is initialized in 1960 ($t = 0$). The results are reported in Appendix Table C.3, where I report five model specifications to test the model’s sensitivity. The main differences across the specifications are the endogenous productivity spillovers of skill and size and the endogenous amenity supply. The first specification includes only the endogenous spillover effect of skill. The second adds the spillover effects of population size on top of the skill spillover. The third adds endogenous amenities but removes the spillover effects of population size. The fourth reports a “classic” model with no production spillovers and no endogenous amenities. Finally, the fifth and last specification reports the full model used for the rest of the analysis below. Overall, the estimates are consistent with the literature. Panel A reports the labor-demand parameters: the CES curvature parameter ρ , the spillover effects of skill (γ^H, γ^L) , and the spillover effects of population size (ϕ^H, ϕ^L) . To interpret ρ , note that the elasticity of substitution between highly skilled and less skilled labor is $\sigma^{HL} = 1/(1 - \rho)$, so ρ is not the inverse elasticity. The implied substitution elasticities, computed from the displayed ρ entries, are approximately 2.50 in column A, 4.00 in column B, 2.13 in column C, 3.23 in column D, and 3.03 in column E. Overall, the estimates suggest that the implied elasticity of substitution is of a similar order of magnitude across specifications, with the “classic” specification lying within the range spanned by the models with production spillovers rather than uniformly below them. The estimates of γ^H are positive and statistically significant in all specifications that include this parameter. In the specifications with no spillover effects of size (columns A and C), a 1% increase in the local skill ratio raises the highly skilled productivity shifter by

0.72%, holding city size and other inputs fixed. Instead, in columns B and E, which include spillover effects of size, the corresponding partial productivity elasticities are 0.55 and 0.53, respectively. This internal comparison suggests that, when city size is omitted, the skill-ratio coefficient absorbs part of the positive association between city size and highly skilled wages. My estimates of γ^H are also higher than in [Diamond \(2016\)](#) and [Moretti \(2004\)](#), which find estimates of 0.322 and 0.16, respectively. Overall, the estimated productivity spillover from the local skill ratio is positive for highly skilled workers. The corresponding estimates for less skilled workers are not statistically significant. The displayed estimates of γ^L are between 0.06 and 0.12, but they are not statistically significant in any specification. The estimates of the spillover effects of size, ϕ^H , for highly skilled workers are 0.27 and 0.25 in columns B and E, respectively. These coefficients should be interpreted as partial elasticities of the local productivity shifter with respect to city size, holding the skill ratio and own labor inputs fixed, rather than as total wage elasticities in general equilibrium. This effect is smaller than the skill-ratio spillover but it is still non-negligible. For less skilled workers, the corresponding estimates in columns B and E are 0.21 and 0.21, but they are not statistically significant. This contrasts with [Baum-Snow and Pavan \(2013\)](#), who estimate that at least 23% of the overall increase in the variance of log hourly wages in the U.S. from 1979 to 2007 is explained by the more rapid growth in the variance of log wages in larger locations relative to smaller locations. Overall, the production-spillover estimates are positive and statistically significant for highly skilled workers, but not for less skilled workers. The estimates also suggest that the spillover on the skill ratio is quantitatively more important than the size spillover.

Panel B reports estimates of the labor-supply parameters, specifically the Stone–Geary marginal expenditure share on tradables, θ , and the preference for endogenous amenities, γ^p . In columns A and B, the displayed estimates of θ are 0.62 and 0.61. Under Stone–Geary preferences, this means that about 60% of supernumerary income is allocated to tradables and about 40% to housing. Because the subsistence housing requirement $\bar{O}$ must be met first, these are marginal rather than average expenditure shares. In columns C–E, the displayed estimates range from 0.42 to 0.43. Specifications C and E, which include endogenous amenities, imply a lower marginal tradables share out of supernumerary income (and a higher marginal housing share) than specifications A and B, which exclude endogenous amenities. The second parameter of interest on the labor supply side is γ^p , estimated only in specifications C and E. In those two specifications, the estimates are very similar. Interpreted as a partial elasticity in the location-choice system, column E implies that a 1% increase in the local skill ratio raises the exponentiated location-choice index for both skill groups by about 0.93%, holding other determinants fixed. This value is lower than that in [Diamond \(2016\)](#) and [Albouy \(2012\)](#). One explanation could be that my estimates relate to a longer period. These estimates show that

workers generally prefer cities with higher wages, lower rents, and higher college shares.

Estimation and Robustness: Speed of Spatial Diffusion. Panel C of Appendix Table C.3 reports the persistence estimates underlying ϱ_H and ϱ_L in equation (19). I estimate the two skill-specific persistence parameters separately and test their equality rather than imposing a common coefficient. The estimating equation uses the positive productivity level, the native-decadal transition, and the unweighted arithmetic global city mean in levels within each skill and sector, exactly as in the simulator. The training sample comprises the decadal transitions from 1960 through 1990; subsequent transitions are excluded from estimation. Sector–transition fixed effects absorb common changes, and standard errors are clustered by city. The estimated persistence parameters are $\hat{\varrho}_H = 0.378$ (s.e. 0.036) and $\hat{\varrho}_L = 0.703$ (s.e. 0.034). An equality test rejects a common persistence parameter ($\chi^2(1) = 63.47$, $p < 0.001$). The remaining city–sector–skill change is recorded in the realized innovation ε_{kdjt} and is held fixed rather than attributed mechanically to diffusion.

Productivity and amenity inversion. The quantitative baseline uses a three-step inversion on the common positive-support city–sector–skill panel. First, at each observed date, I use firms’ sectoral first-order conditions and goods-market clearing to recover the effective productivity terms η_{kdjt}^{obs} that are consistent with observed sectoral allocations. I then divide these recovered effective productivity terms by the structural composition, city-size, and SBTC-exposure factors in equation (18). The remainder is the positive baseline productivity level $\exp(\xi_{kdjt})$ used in equation (19).

Second, I recover a city–skill–date productivity residual b_{kjt} , common across sectors, to match observed skill-specific centered wages at observed allocations. I normalize the residual to have zero mean across cities, $\sum_{j=1}^N b_{kjt} = 0$, so that it captures the remaining spatial variation in wages while leaving the national skill–date wage component governed by the baseline specification. In the quantitative implementation, I add this residual to log productivity, replacing $\exp(\xi_{kdjt})$ in equation (18) with $\exp(\xi_{kdjt} + b_{kjt})$. Aggregate wage growth and premium levels are therefore not matched exactly.

Third, holding fixed the time-invariant destination component $m_{k2}(j)$ of the bilateral migration cost, I invert the amenity vector once for each observed skill–date cell. Define $s_{kjt} = E_{kjt}/E_{kt}^{tot}$, where $E_{Hjt} = H_{jt}$ and $E_{Ljt} = L_{jt}$, and let $\tilde{\delta}_{kjt} \equiv \delta_{kjt} - A_{kjt}$, using the preference index in equation (6). The closed-form share inversion is

$$A_{kjt} = q_{kjt} - \frac{1}{N} \sum_{r=1}^N q_{krt}, \quad q_{kjt} = \log s_{kjt} - \tilde{\delta}_{kjt} + \log m_{k2}(j). \quad (33)$$

This normalization gives $\sum_j A_{kjt} = 0$ and reproduces the observed location shares exactly at each of the six dates. The realized productivity innovations, productivity residuals, and amenities are then held fixed across counterfactuals. Appendix Subsection C.6 provides the full equations, normalizations, and numerical sequence.

Validation of Amenities and Destination Components of Bilateral Migration

Costs. Under the decomposition above, the recovered A_{kjt} captures date-specific residual attractiveness, while $m_{k2}(j)$ is the time-invariant destination component of the bilateral migration cost. Because a common multiplicative rescaling of this component leaves location shares unchanged, I use the scale normalization $\min_j m_{k2}(j) = 1$ and interpret $m_{k2}(j)$ as a cross-city measure of persistent entry frictions. Appendix Section C.4 compares both objects with independent metropolitan proxies and reports their rankings over the 70 locations used by the quantitative model.²¹ I assess the economic interpretation of the recovered amenities and migration-cost components by comparing them with observed city characteristics.

Table C.7 shows that New York, Los Angeles, and Chicago occupy the first three positions for both skill groups in the 2010 amenity ranking. The remaining top-ten positions mix large Sun Belt, coastal, and northern metropolitan areas, while Flint, Worcester, South Bend, and Trenton appear among the bottom positions for both groups. Table C.8 shows that New York, Los Angeles, Chicago, and Philadelphia have the four largest time-invariant destination components of bilateral migration costs for both skill groups. These rankings and the proxy correlations in Appendix Tables C.5 and C.6 provide descriptive validation evidence. In particular, only six amenity-proxy rows overlap the 70-city model universe and their row-specific complete-pair counts range from three to six.

Quantitative SBTC scaling parameters. The quantitative exercises use the baseline SBTC exposure path. Alternative exposure specifications rescale this baseline using skill-specific coefficients that interpolate linearly between 1960 and 2010 endpoint values. These endpoint pairs are $(-0.04, 0.01)$ for highly skilled workers and $(0.03, -0.04)$ for less skilled workers.

Appendix Table C.4 compares aggregate wage growth and the skill premium in the model with the observed wage panel for the same 70 MSAs. Both columns use skill-specific employment weights and the same aggregation formula. Premium growth is the difference between the two wage-growth rates; the 2010 premium uses the observed 1980 premium in this sample as the common starting point.

²¹Appendix Tables C.5 and C.6 report the external-proxy correlations, while Tables C.7 and C.8 report the corresponding model-universe rankings. Appendix Section C.6 separately documents the high-dimensional productivity inputs.

For the alternative exposure specifications, write the endpoint coefficients as $\lambda_k^{t_0}$ and $\lambda_k^{t_1}$. Their interpolation is

$$\Lambda_{kt} = (1 - g_t)\lambda_k^{t_0} + g_t\lambda_k^{t_1}, \quad k \in \{H, L\},$$

where $g_t = (t - t_0)/(t_1 - t_0)$, with $t_0 < t_1$ denoting the initial and terminal dates and $t \in [t_0, t_1]$.

These endpoint coefficients are distinct from the nuisance coefficients on $\Delta \ln S_{Hjt}$ and $\Delta \ln S_{Ljt}$ in the first-step wage equations. The latter help recover productivity residuals from cross-city variation; the former govern the rescaling used for alternative exposure specifications. This specification does not affect the identification of agglomeration and amenity parameters, which is driven by cross-city differential equilibrium responses to exogenous demand shifters (and their interaction with housing supply constraints), rather than by the scale normalization of the SBTC exposure indices.

The baseline closely reproduces aggregate wage growth for highly skilled workers. For less skilled workers, the model generates a smaller wage decline and therefore a smaller increase in the skill premium than in the observed panel. This within-sample comparison assesses aggregate outcomes alongside the reconstruction of cross-city wage profiles.

Inversion of the realized productivity path. The productivities recovered at the six observed dates imply the realized innovations between consecutive dates:

$$\varepsilon_{kdjt} = \xi_{kdjt} - \varrho_k \xi_{kdj,t-1} - (1 - \varrho_k) \log \left(\frac{1}{N} \sum_s \exp(\xi_{kds,t-1}) \right).$$

Thus the estimated persistence parameters govern the systematic convergence component, while ε_{kdjt} records the realized change needed to reconcile the observed productivity geography with that transition. The entire realized-innovation path is fixed before any mechanism shutdown.

Trade Costs I parameterize domestic bilateral iceberg trade costs as a monotone function of bilateral distance and impose the iceberg restriction $\tau_{djj'} \geq 1$ such that $\tau_{djj'} = \min\{\max(\text{distance}_{jj'}^\kappa, 1), \bar{\tau}\}$, with $\kappa > 0$ and an upper cap $\bar{\tau}$ to avoid extreme tails in the distance matrix. In the baseline, I set $\kappa = 0.5$ and $\bar{\tau} = 50$ (Appendix Section C.1). This parameterization is consistent with the gravity equation $X_{jj'}^d \propto (\tau_{jj'}^d)^{1-\sigma_d}$ and $\sigma_d > 1$. When mapping empirical distance elasticities from the trade literature (e.g., [Eckert 2019](#)) into this iceberg representation, the relevant object is the implied cost elasticity $\kappa_d = -\beta_d/(\sigma_d - 1)$, where β_d is the estimated distance coefficient in trade flows.

Rest of the World (ROW). Following [Eckert \(2019\)](#), I introduce the Rest of the World (ROW) as an additional region in the trade matrix. The trade block closes domestic and international goods markets jointly within the spatial equilibrium. Trade costs and openness determine bilateral expenditure shares, sectoral producer prices, destination price indices, real wages, and migration incentives, while ROW expenditure adjusts so that aggregate exports equal aggregate imports. The ROW trades with all domestic regions under calibrated trade costs $\tau_{j,ROW,d}$, chosen so that the model reproduces observed U.S. import shares from the ROW separately for goods (18%) and services (6%). ROW sectoral producer prices are fixed exogenously and normalized to one. Holding these prices fixed makes ROW an infinitely elastic supplier at the normalized world prices. There is no international migration to or from the ROW, so labor mobility remains confined within domestic regions. This addition leaves the domestic Armington trade structure unchanged but allows the model to quantify how external openness affects regional convergence through local prices and expenditure shares (Appendix Section [D.3](#)).

Housing Subsistence and Elasticity To complete the calibration of the model and compute its equilibrium, I fix the remaining housing-sector parameters outside the GMM step. These values are reported in Appendix Section [C](#). To include housing in the model with non-homothetic preferences, I set the subsistence level of housing to $\bar{O} = 0.25$, following [Ganong and Shoag \(2017\)](#). For the housing-supply elasticity, I do not estimate μ structurally in the baseline. I use one common housing-supply elasticity across cities, equal to 0.40 through 1980 and declining by 0.05 per decade thereafter. Hence the baseline does not use location-specific housing-supply elasticities. These values are chosen so that the model generates a broadly realistic positive relationship between local prices and local incomes, while remaining in the range used in the housing literature, including [Ganong and Shoag \(2017\)](#).

4.1 Model Solution and Spatial Moments

Given the estimated parameters and inverted inputs from Section [4](#), I solve the dynamic spatial equilibrium recursively as a sequence of competitive static equilibria linked by equation [\(19\)](#). Starting from the recovered 1960 stock, I carry productivity forward using the separately estimated persistence parameters and the recovered productivity innovations. At each date I combine this stock with the fixed SBTC exposure mapping, the mean-zero productivity residual, and the endogenous agglomeration terms, then solve jointly for wages, rents, prices, sectoral labor demand, and skill-specific location shares under goods-market clearing. The six-date amenity values are supplied from the single baseline inversion and remain frozen across counterfactuals.

The equilibrium for each t is obtained by fixed-point iteration in $\{W_{kjt}, R_{jt}, P_{dj}, H_{jt}, L_{jt}\}$ until changes across iterations fall below 10^{-6} . The resulting sequence $\{W_{kjt}, R_{jt}, P_{jt}, H_{jt}, L_{jt}\}$ yields the model’s dynamic path at the observed decadal dates. At the estimated parameter values, the algorithm converges from the tested starting values.²² Additional computational details and the distinction between sufficient analytical conditions and numerical convergence are provided in Appendix Subsection C.1.

For the model–data comparison, I estimate equation (1) over overlapping 20-year windows advanced in ten-year increments on the common sample of cities. These are the skill-specific wage-convergence outcomes documented in Fact 2. Appendix Figure C.2 reports the resulting convergence paths on the common observed support. The fit of the convergence coefficients follows from matching the centered wage profiles. The left (right) panel compares β^H (β^L): the reconstruction preserves the observed shift from high-skill convergence to divergence and continued low-skill convergence. The counterfactuals below hold fixed the realized productivity innovations, productivity residuals, amenities, and all components not named in the exercise.

Secular Migration, Skill-Ratio Convergence, and Sorting. Appendix Table C.10 compares the baseline model with the observed input panel using the same reporting operator on the common city and date support. The first two columns report the change in average annual secular migration from 1961–1980 to 1990–2010. The data calculation interpolates city employment log-linearly between decadal observations. The third column compares the 1960–1980 and 1990–2010 skill-ratio convergence coefficients: annualized 20-year growth in the log skill ratio is regressed on its initial log level, weighted by initial total employment. This is the model-support counterpart to Fact 5 (Section 2.5). The fourth column reports the 1980–2010 change in the unweighted correlation between the log skill premium and the log skill ratio. This correlation rises by 0.20 in the model and 0.19 in the data, indicating stronger sorting by skill abundance as in Fact 3 (Section 2.3). The model closely tracks the observed changes without matching them exactly.

5 The Role of SBTC and Agglomeration in the End of Regional Convergence

After validating and comparing the model with the data, I use the model to quantify which mechanisms account for the post-1980 slowdown in regional convergence. I do so through a sequence of counterfactual “shutdown” exercises that proceed sequentially: starting from the

²²This numerical behavior is not a proof that the calibrated mapping satisfies the contraction condition globally or that the equilibrium is globally unique.

baseline economy, I remove one model component at a time and recompute the equilibrium transition path. Importantly, these counterfactuals are *not* recalibrated; parameters disciplined in the baseline are held fixed across shutdowns so that differences in convergence moments reflect the mechanism being removed rather than refitting a new economy.

My main comparison uses one pre-specified reporting convention for every model path. For each skill group $k \in \{H, L\}$, I compute the raw 20-year convergence coefficient using endpoint-period skill-employment weights. There is no temporal smoothing. I summarize the “end of convergence” by $\Delta\beta^k \equiv \beta_{2010}^k - \beta_{1980}^k$.

Table 1 reports the convergence moments in Panel A and the associated migration, skill-divergence, and sorting moments in Panel B. In the model, highly skilled wages move from

Table 1: Counterfactual Changes in Convergence and Related Spatial Moments in the model

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
No Agglomeration	-1.90	0.10	2.00 pp (105.52%)	-1.55	-1.53	0.02 pp (1.25%)
No SBTC	-1.90	0.97	2.87 pp (151.13%)	-1.55	-2.08	-0.53 pp (-34.17%)
No Agg. $\times$ No SBTC	-1.90	-0.03	1.87 pp (98.43%)	-1.55	-1.62	-0.07 pp (-4.81%)
Housing Tightening	-1.90	0.78	2.68 pp (141.24%)	-1.55	-1.78	-0.23 pp (-15.01%)
Free Migration	-1.90	1.61	3.51 pp (184.99%)	-1.55	-1.10	0.45 pp (28.95%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
No Agglomeration	-0.04 pp		0.07 pp	-1.44 pp		-0.67
No SBTC	-0.61 pp		-0.39 pp	0.38 pp		0.30
No Agg. $\times$ No SBTC	-0.03 pp		0.08 pp	-1.45 pp		-0.60
Housing Tightening	-0.54 pp		-0.36 pp	0.26 pp		0.15
Free Migration	-0.69 pp		-0.43 pp	0.34 pp		0.45

Note: Table 1 reports the counterfactual results in two panels. Panel A reports β^H and β^L at 1980 and 2010 and their changes, expressed in percentage points with percent changes in parentheses. Panel B reports changes in the model’s secular migration indices, $\Delta Mig.^H$ and $\Delta Mig.^L$, the skill-ratio convergence coefficient ($\Delta\beta^{HL}$), and the correlation between the log skill premium and the log skill ratio ($\Delta\delta$). The secular migration columns compare the 1990–2010 average with the 1960–1980 average. The remaining columns compare 2010 to 1980. Secular migration and $\Delta\beta^{HL}$ changes are reported in percentage points, while $\Delta\delta$ is reported in levels. Rows correspond to Model, No Agglomeration, No SBTC, No Agglomeration $\times$ No SBTC, Housing Tightening, and Free Migration.

convergence in the early window to divergence in the late window, while less-skilled wages remain convergent and become modestly more so. The post-1980 shifts are $\Delta\beta_{2010-1980}^H = 2.93$ pp (154.74%) and $\Delta\beta_{2010-1980}^L = -0.36$ pp (-23.28%). Panel B shows that the same model reconstruction displays declines in average secular migration between 1960–1980 and 1990–2010 for both groups, with $\Delta Mig.^H = -0.61$ pp and $\Delta Mig.^L = -0.37$ pp. It also displays a

1980–2010 increase in the skill-ratio divergence moment of $\Delta\beta^{HL} = 0.37$ pp and a change in the correlation between the log skill premium and the log skill ratio of $\Delta\delta = 0.20$. These outcomes summarize the model’s secular migration, skill-divergence, and sorting margins.

Full agglomeration shutdown. The “No Agglomeration” counterfactual applies the post-1980 schedule described above to the full set of local productivity and utility/amenity spillovers while leaving the pre-1980 history unchanged. For the city-size term, the schedule preserves the 1980 stock already embodied in local productivity and removes only the post-1980 ϕ^H flow; it is not a rewrite of the initial productivity geography. The post-1980 changes then become $\Delta\beta^H = 2.00$ pp and $\Delta\beta^L = 0.02$ pp. Thus, the highly skilled reversal remains but is materially attenuated. For less-skilled wages, the model’s modest strengthening of convergence is replaced by an almost unchanged coefficient. The decline in secular migration is much smaller for highly skilled workers ($\Delta Mig.^H = -0.04$ pp), while the less skilled decline becomes a small increase ($\Delta Mig.^L = 0.07$ pp). At the same time, $\Delta\beta^{HL}$ falls from 0.37 pp in the baseline to -1.44 pp, and $\Delta\delta$ changes from 0.20 to -0.67 . The full spillover block therefore has substantial effects on both the convergence coefficients and the related spatial moments in this calibrated economy.

No SBTC. The No SBTC counterfactual makes two changes after 1980. First, I gradually reduce highly skilled workers’ productivity relative to the baseline by the same proportion across cities. Second, I hold each city’s relative exposure to technology favoring highly skilled over less skilled workers at its 1980 level. The experiment therefore combines a nationwide reduction in the productivity advantage of highly skilled workers with a freeze in local technological exposure. It preserves the economy’s pre-1980 history and leaves other model inputs unchanged, while allowing wages, employment, and workers’ location choices to respond.²³ Under this change, $\Delta\beta^H = 2.87$ pp and $\Delta\beta^L = -0.53$ pp. The highly skilled shift is only modestly smaller than in the model, while less-skilled convergence strengthens more. Secular migration changes remain close to the model: $\Delta Mig.^H = -0.61$ pp, $\Delta Mig.^L = -0.39$ pp, $\Delta\beta^{HL} = 0.38$ pp, and $\Delta\delta = 0.30$. The sorting response is larger than in the baseline, but the standalone effect on high-skill convergence remains small.

No Agglomeration $\times$ No SBTC. The “No Agglomeration $\times$ No SBTC” row applies exactly the full productivity and utility/amenity spillover shutdown above together with the

²³The national multiplier declines log-linearly from one in 1980 to 0.75 in 2010. Its slope is chosen to eliminate 1980–2010 aggregate skill-premium growth under the national adjustment alone, before freezing local exposure, and is then held fixed in both No SBTC and the joint agglomeration shutdown. The local freeze preserves the baseline geometric mean of the two skill-specific exposure factors at each city–sector–date.

same national adjustment and local exposure freeze. Neither is re-estimated nor retargeted in the joint economy. The joint outcomes lie close to the full No Agglomeration row, with a slightly smaller high-skill shift, a small negative low-skill shift, and nearly unchanged Panel B moments. For highly skilled workers, removing both mechanisms reduces the model’s convergence reversal by about 36 percent. This reduction exceeds the sum of their individual effects and is the largest among the main mechanism counterfactuals. For less-skilled wages, the joint change in the convergence coefficient exceeds the sum of the separate changes.

Targeted housing tightening. The “Housing Tightening” counterfactual uses a fixed treated set based on 1980 data: the union of metros in the top 5% of the rent distribution and the top 10 metros by highly skilled wages. It lowers only the treated metros’ housing-supply elasticity endpoint, from the baseline $\mu_2 = 0.25$ to approximately 0.08 subject to a 0.05 floor, while leaving SBTC, agglomeration, migration costs, and trade openness at baseline. The resulting changes are $\Delta\beta^H = 2.68$ pp and $\Delta\beta^L = -0.23$ pp. Relative to the model, the highly skilled reversal is attenuated, and the strengthening of less-skilled convergence is also somewhat attenuated. The secular migration changes, -0.54 pp and -0.36 pp, the skill-divergence moment, 0.26 pp, and the sorting moment, 0.15 , remain close to their baseline values. Targeted housing tightening therefore does not reproduce the broader bundle of changes generated by the full spillover shutdown.

Free migration. The “Free Migration” counterfactual preserves the estimated destination components of the bilateral migration costs through 1980 and sets them to one at decennial model dates from 1990 through 2010. The exercise compares static equilibria at the reported dates, using the corresponding destination components in the location-choice system. Under this change, the high-skill convergence shift is $\Delta\beta^H = 3.51$ pp and the less-skilled shift is $\Delta\beta^L = 0.45$ pp. The skill-ratio convergence coefficient changes by $\Delta\beta^{HL} = 0.34$ pp and the sorting statistic changes by $\Delta\delta = 0.45$. The change in average secular migration is $\Delta Mig.^H = -0.69$ pp for highly skilled workers and $\Delta Mig.^L = -0.43$ pp for less skilled workers. Removing the post-1980 destination components of the bilateral migration costs does not restore high-skill convergence.

Overall, the joint shutdown produces the largest attenuation of the baseline shift toward highly skilled wage divergence. Among the individual shutdowns, No Agglomeration produces the largest attenuation of that shift. The No SBTC row remains close to baseline, targeted housing tightening changes the convergence coefficients without materially changing the broader Panel B moments, and free migration does not restore convergence. These comparisons quantify the effects of the specified interventions within the calibrated model. Table 1 therefore

provides the main decomposition in the text; complementary mechanisms (trade openness, human capital accumulation, physical capital, and no-diffusion variants) are evaluated in Section 6 and Appendix Section D.

6 Robustness

This section evaluates whether three prominent alternative channels could account for the post-1980 change in regional β -convergence documented in the data and reconstructed by the quantitative model in Section 5: *physical capital and transition dynamics*, *human capital accumulation*, and *trade openness*. The goal is not to rule out that these forces matter for regional outcomes, but to assess whether they can replicate the *timing* and *skill asymmetry* of the convergence reversal—namely that the breakdown in convergence is concentrated among highly skilled wages.

Physical capital and transition dynamics. The baseline model solves a sequence of within-period spatial equilibria and therefore abstracts from the potentially slow adjustment of predetermined physical capital. If capital-induced transition dynamics were the main driver of the post-1980 patterns, then (i) the breakdown in convergence should be visible at broader geographic levels and (ii) it should not be as sharply concentrated among highly skilled wages. Appendix Section D.1 addresses this concern directly. First, Appendix Section D.1.1 extends the model to include *location-specific physical capital* and forward-looking investment, as in the recent dynamic spatial economics literature (e.g., Kleinman, Liu and Redding 2023), generating additional transition dynamics absent from the baseline. Appendix Table D.2 shows that this extension is a material sensitivity. Because capital is present from 1960, its early-window coefficients need not equal the baseline. More importantly, the capital economy produces stronger convergence for highly skilled wages and weaker convergence for less skilled wages. This is the opposite of the empirical skill asymmetry that motivates the paper. Thus, in the model, physical capital increases convergence rather than generating divergence, especially for highly skilled wages. Second, the evidence shows that at the state level the post-1980 convergence slowdown for highly skilled wages is much weaker than at the MSA level (Appendix Table D.1), consistent with a mechanism that operates primarily through local agglomeration and sorting rather than through aggregate capital reallocation alone. Finally, Appendix Section D.1.2 evaluates a closely related alternative mechanism—endogenous innovation via capital–skill complementarity. It shows that it delivers implications for the skill premium relative to skill supply that are inconsistent with the empirical patterns motivating the SBTC channel.

Human capital accumulation. A second concern is that the post-1980 divergence could be driven by heterogeneous *human-capital accumulation* across locations (for example, differential growth in college attainment across cities) rather than by the interaction of SBTC and local agglomeration emphasized by the model. Empirically, post-1980 wage-growth regressions with controls show that the initial cross-city skill distribution receives the larger share of the two controls’ incremental explanatory power. Appendix Table D.3 reports a Shapley allocation of their joint contribution to R^2 , averaged across the possible orderings. To assess the quantitative importance of aggregate skill upgrading in the model, Appendix Section D.2 implements a counterfactual that holds the economy-wide skill shares fixed at their 1980 levels from 1980 onward, shutting down the secular rise in the college share while keeping all other primitives unchanged. Appendix Table D.4 shows that this counterfactual changes the predicted convergence shifts differently by skill. In particular, the highly skilled convergence change falls from $\Delta\beta_{2010-1980}^H = 2.93$ pp in the baseline to 2.50 pp under No H/L Growth, while the less skilled change moves from -0.36 pp to -0.73 pp. Fixing aggregate skill growth attenuates the high-skill reversal and makes the low-skill change more negative, strengthening low-skill convergence. At the same time, the aggregate skill-ratio convergence moment only falls from 0.37 pp to 0.27 pp, and the secular migration declines are nearly unchanged: $\Delta Mig.^H = -0.61$ pp under No H/L Growth versus -0.61 pp in the baseline, and $\Delta Mig.^L = -0.38$ pp versus -0.37 pp. Thus, aggregate education growth is quantitatively important for the magnitudes of the convergence break, but the baseline mechanism is not simply proxying for aggregate education growth.

Trade openness. A third concern is that international trade could drive the post-1980 divergence by shifting local labor demand in a way correlated with initial skill composition. In the data, controlling for exposure to the China trade shock leaves the convergence patterns essentially unchanged (Appendix Table D.5). In the quantitative model, I conduct a separate counterfactual exercise by shutting down trade openness to the Rest of the World (ROW), effectively imposing trade autarky vis-à-vis ROW goods trade (see Appendix Section D). The resulting counterfactual still displays a sizable post-1980 slowdown in highly skilled convergence and differs only modestly from the model. In particular, Appendix Table D.6 shows that $\Delta\beta_{(90-10)-(60-80)}^H$ falls from 2.93 pp in the baseline to 2.91 pp under No ROW, while $\Delta\beta_{(90-10)-(60-80)}^L$ changes from -0.36 pp to -0.37 pp. In other words, shutting down ROW trade changes the highly skilled divergence margin by -0.02 pp. The aggregate skill-ratio convergence moment remains large at 0.36 pp, compared with 0.37 pp in the baseline, and the secular migration declines also remain substantial ($\Delta Mig.^H = -0.61$ pp and $\Delta Mig.^L = -0.37$ pp, versus -0.61 pp and -0.37 pp in the baseline). Removing ROW trade changes the

skill-specific convergence reversal only slightly in this comparison.

In addition to these three channels, Appendix Section D reports a broader set of empirical checks and alternative mechanisms, such as changes in sectoral composition and IT intensity (Appendix Table D.8) and institutional changes captured by unionization and right-to-work laws (Appendix Table D.9 and Figure D.3). It also documents that the main convergence estimates are robust to alternative wage measures and specifications (Online Appendix Table OA.1).

7 Conclusions and Future Directions

This paper shows that the decline in regional β -convergence among MSAs observed after 1980 is primarily due to the decline in regional β -convergence among highly skilled workers. Thus, any account of β -convergence and its decline must distinguish between skill groups. Motivated by this observation, I develop a quantitative spatial equilibrium model with technology diffusion, agglomeration and SBTC. I bring the model to the data and use counterfactual experiments to quantify the contribution of these mechanisms to regional wage convergence.

Among the individual mechanism counterfactuals in Table 1, removing local agglomeration produces the largest reduction in the high-skill convergence reversal. Combining its shutdown with the No SBTC path lowers the post-1980 highly skilled convergence shift by about 36 percent, the largest reduction among the main mechanism shutdowns, with a more-than-additive response. These mechanisms also account for much of the model’s decline in secular migration: their joint removal largely eliminates that decline.

Understanding what stopped income convergence across U.S. regions and increased income inequality for different skill levels might have important policy implications, especially for the regions that cannot grow like richer ones. Sustaining growth in the richest MSAs while arresting decline in poorer MSAs is a fundamental challenge for policymakers.

This paper opens several avenues for future research. One natural extension is to rationalize the patterns of regional inequality observed across countries. The framework developed here is sufficiently flexible to accommodate a range of applications, including cross-country and within-country analyses. Recent evidence shows that the decline in regional convergence is a global phenomenon affecting a large share of countries worldwide. Although poorer countries continue to grow faster than richer ones, much of this growth is concentrated in a few regions, leaving others behind (Chatterjee et al., 2025). These findings underscore the need to further understand the evolving nature of regional convergence both in the United States and globally.

References

- Acemoglu, Daron, and David Autor.** 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” , ed. O. Ashenfelter and D. Card Vol. 4 of *Handbook of Labor Economics*, Chapter 12, 1043–1171. Elsevier.
- Adão, Rodrigo, Michal Kolesár, and Eduardo Morales.** 2019. “Shift-Share Designs: Theory and Inference.” *The Quarterly Journal of Economics*, 134(4): 1949–2010.
- Albouy, David.** 2012. “Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas.” National Bureau of Economic Research Working Paper 14472.
- Autor, David H.** 2019. “Work of the Past, Work of the Future.” Vol. 109, 1–32, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Autor, David H., and David Dorn.** 2013. “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market.” *American Economic Review*, 103(5): 1553–1597.
- Barro, Robert J., and Xavier Sala-i Martin.** 1992. “Convergence.” *Journal of Political Economy*, 100(2): 223–251.
- Barro, Robert J., and Xavier Sala-i Martin.** 1997. “Technological Diffusion, Convergence, and Growth.” *Journal of Economic Growth*, 2(1): 1–26.
- Baumol, William.** 1986. “Productivity Growth, Convergence, and Welfare: What the Long-run Data Show.” *American Economic Review*, 76(5): 1072–85.
- Baum-Snow, Nathaniel, and Ronni Pavan.** 2013. “Inequality and City Size.” *The Review of Economics and Statistics*, 95(5): 1535–1548.
- Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan.** 2018. “Why Has Urban Inequality Increased?” *American Economic Journal: Applied Economics*, 10(4): 1–42.
- Beaudry, Paul, Mark Doms, and Ethan Lewis.** 2010. “Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas.” *Journal of Political Economy*, 118(5): 988 – 1036.
- Bernard, Andrew B, and Charles I Jones.** 1996. “Productivity and Convergence across U.S. States and Industries.” *Empirical Economics*, 21(1): 113–35.
- Berry, Christopher R., and Edward L. Glaeser.** 2005. “The Divergence of Human Capital Levels Across Cities.” *Papers in Regional Science*, 84(3): 407–444.

- Berry, Steven, James Levinsohn, and Ariel Pakes.** 2004. “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market.” *Journal of Political Economy*, 112(1): 68–105.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel.** 2022. “Quasi-Experimental Shift-Share Research Designs.” *The Review of Economic Studies*, 89(1): 181–213.
- Bound, John, and George Johnson.** 1992. “Changes in the Structure of Wages in the 1980’s: An Evaluation of Alternative Explanations.” *American Economic Review*, 82(3): 371–92.
- Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro.** 2019. “Trade and labor market dynamics: General equilibrium analysis of the china trade shock.” *Econometrica*, 87(3): 741–835.
- Card, David, and John E. DiNardo.** 2002. “Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles.” *Journal of Labor Economics*, 20(4): 733–783.
- Caselli, Francesco, and Wilbur John Coleman.** 2001. “The U.S. Structural Transformation and Regional Convergence: A Reinterpretation.” *Journal of Political Economy*, 109(3): 584–616.
- Chatterjee, Shoumitro, Elisa Giannone, Tatjana Kleineberg, Kan Kuno, and Luca Looser.** 2025. “Unequal Global Convergence.” Working paper.
- Comin, Diego A., Mikhail Dmitriev, and Esteban Rossi-Hansberg.** 2012. “The Spatial Diffusion of Technology.” National Bureau of Economic Research Working Paper 18534.
- Comin, Diego, and Martí Mestieri.** 2018. “If technology has arrived everywhere, why has income diverged?” *American Economic Journal: Macroeconomics*, 10(3): 137–178.
- Davis, Donald R., and Jonathan I Dingel.** 2019. “A spatial knowledge economy.” *American Economic Review*, 109(1): 153–170.
- Davis, Donald R., and Jonathan I. Dingel.** 2020. “The Comparative Advantage of Cities.” *Journal of International Economics*, 123: 103291.
- Desmet, Klaus, and Esteban Rossi-Hansberg.** 2014. “Spatial Development.” *American Economic Review*, 104(4): 1211–43.

- Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg.** 2018. “The Geography of Development.” *Journal of Political Economy*, 126(3): 903–983.
- Diamond, Rebecca.** 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review*, 106(3): 479–524.
- Eckert, Fabian.** 2019. “Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy.” Society for Economic Dynamics Meeting Paper 307.
- Eckert, Fabian, and Michael Peters.** 2022. “Spatial Structural Change.” National Bureau of Economic Research NBER Working Paper 30489.
- Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny.** 2014. “Spatial Sorting.” *Journal of Political Economy*, 122(3): 554–620.
- Fajgelbaum, Pablo D., and Cecile Gaubert.** 2020. “Optimal Spatial Policies, Geography, and Sorting.” *The Quarterly Journal of Economics*, 135(2): 959–1036.
- Ganong, Peter, and Daniel Shoag.** 2017. “Why Has Regional Income Convergence in the U.S. Declined?” *Journal of Urban Economics*, 102: 76–90.
- Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer.** 2014. “Growth in Regions.” *Journal of Economic Growth*, 19(3): 259–309.
- Glaeser, Edward L., Joseph Gyourko, Eduardo Morales, and Charles G. Nathanson.** 2014. “Housing Dynamics: An Urban Approach.” *Journal of Urban Economics*, 81: 45–56.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift.** 2020. “Bartik Instruments: What, When, Why, and How.” *American Economic Review*, 110(8): 2586–2624.
- Hsieh, Chang-Tai, and Enrico Moretti.** 2019. “Housing Constraints and Spatial Misallocation.” *American Economic Journal: Macroeconomics*, 11(2): 1–39.
- Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow.** 2019. “The Allocation of Talent and U.S. Economic Growth.” *Econometrica*, 87(5): 1439–1474.
- Katz, Lawrence F., and Kevin M. Murphy.** 1992. “Changes in Relative Wages, 1963–1987: Supply and Demand Factors.” *The Quarterly Journal of Economics*, 107(1): 35–78.
- Kleinman, Benny, Ernest Liu, and Stephen J. Redding.** 2023. “Dynamic Spatial General Equilibrium.” *Econometrica*, 91(2): 385–424.

- Levy, Frank, and Richard J Murnane.** 1992. “U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations.” *Journal of Economic Literature*, 30(3): 1333–81.
- Lyon, Spencer G., and Michael E. Waugh.** 2019. “Quantifying the Losses from International Trade.” Unpublished manuscript, New York University.
- Moretti, Enrico.** 2004. “Human Capital Externalities in Cities.” *Handbook of Regional and Urban Economics*.
- Moretti, Enrico.** 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.
- Nagy, Dávid Krisztián.** 2023. “Hinterlands, City Formation and Growth: Evidence from the U.S. Westward Expansion.” *The Review of Economic Studies*, 90(6): 3238–3281.
- Roback, Jennifer.** 1982. “Wages, Rents, and the Quality of Life.” *Journal of Political Economy*, 90(6): 1257–1278.
- Rosen, Susan.** 1979. “Wages-based Indexes of Urban Quality of Life.” *Current Issues in Urban Economics*.
- Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman.** 2026. “Cognitive Hubs and Spatial Redistribution.” *American Economic Journal: Macroeconomics*, 18(2): 72–111.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek.** 2015. “Integrated Public Use Microdata Series: Version 6.0 [dataset].”
- Suárez Serrato, Juan Carlos, and Owen Zidar.** 2016. “Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms.” *American Economic Review*, 106(9): 2582–2624.

Appendix

Table of Contents

A	Additional Evidence on the Main Empirical Facts	A1
A.1	Data and Definitions	A1
A.2	Tables	A3
B	Theory Appendix	B1
B.1	Definition of Equilibrium	B2
B.2	Existence and conditional uniqueness of equilibrium	B5
B.3	Proof of static existence	B7
C	Additional Details on Model Parametrization	C1
C.1	Description of the Computational Algorithm	C1
C.2	Aggregation of the Static Labor-Demand Block	C3
C.3	Additional Details on Measuring Skill-Biased Technical Change and Core Parameters	C6
C.4	Validation of Amenities and Destination Components of Bilateral Migration Costs	C9
C.5	Alternative Approaches to Migration Costs	C14
C.6	Productivity, Wage, and Amenity Inversion	C14
C.7	Spatial Moments in the model and Data	C19
D	Complementary Explanations	D1
D.1	Physical Capital, Complementarity, and Transition Dynamics	D1
D.1.1	Model with Physical Capital and Transition Dynamics	D2
D.1.2	Model with Endogenous Innovation Rate: Capital-Skill Complementarity	D5
D.2	Skill Accumulation over Time	D6
D.3	International Trade	D7
D.4	The Importance of Local Skill-Biased Technical Change	D9
D.5	Housing Regulation	D12
D.6	Structural Transformation Towards IT Service and Communication	D12
D.7	Right-to-Work Laws and Unions	D19
D.8	Evidence of Spatial Technology Diffusion	D20
D.8.1	Empirical Evidence	D20
D.8.2	Quantitative Comparison Without Technology Diffusion	D21

E	Analytical Illustration: Skill Bias, Agglomeration, and Wage Responses	E1
E.1	Baseline Two-Skill, Two-Period Model	E1
E.2	Extension with Population Agglomeration Shifters	E5
E.3	Time-Varying Skill Bias	E8

A Additional Evidence on the Main Empirical Facts

A.1 Data and Definitions

MSA The unit of geography is the metropolitan statistical area (MSA), defined as “a region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core.” I rank the MSAs by the ratio of highly skilled to less skilled workers. I define “highly skilled” MSAs as those with a concentration of highly skilled workers larger than the national average. The remainder are defined as “less skilled” MSAs. I refer to MSAs as cities in the first part of the paper for a less technical discussion. There are two main reasons why I pick MSAs over states or counties. First, MSAs are the smallest unit of analysis for which I can measure wages by skill group, the numbers of highly skilled workers and less skilled workers, and rent by skill group back to 1940. Second, MSAs are consistent with the mechanism I want to explain in this paper. For instance, agglomeration forces happen in San Francisco, not in California. The Census/ACS sample consistently includes 240 MSAs across the four decades from 1980 to 2010, but coverage is less stable from 1940 to 1970. Following the definitions of metropolitan and micropolitan statistical areas, I try to homogenize the definitions of MSAs over time. However, this is not possible for all cities.¹

Highly Skilled and Less Skilled Workers I follow prior work, such as [Acemoglu and Autor \(2011\)](#), that uses education as a proxy for skill. I then create two groups: “highly skilled” workers are those with at least a four-year bachelor’s degree, while “less skilled” workers are those with less education.

Composition-Adjusted Wages I compute hourly wages at the individual level as annual wages divided by the number of hours worked in the last year. My estimation sample consists of individuals between 21 and 55 years of age who were employed at least 40 weeks per year and were not self-employed. However, for a robustness check, I relax the sample restrictions, and the results are qualitatively unchanged. To conduct this robustness exercise, I construct a demographically adjusted wage measure using regressions estimated separately by year and skill group. In this robustness check, I adjust wages for age, sex, and birthplace-group controls only. 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 an adjusted wage measure based on observables:

¹Most of my analyses are also run at the state level, which eliminates any concern of time comparability. The results of the analysis for aggregate β -convergence are similar between states and MSAs.

$$w_{ijkt} = \gamma_{kt} + \Gamma_{kt}X_{it} + \epsilon_{ijkt}$$

where w_{ijkt} is the log of hourly wages of worker i living in MSA j at time t and belonging to skill group k . I estimate this relationship separately for each year and for workers with and without a college degree, and then reconstruct an adjusted wage by adding the residualized wage back to the corresponding year-by-skill mean. I collect these characteristics in X_{it} . The controls in X_{it} include age-bin dummies, one dummy for sex, and birthplace-group dummies. The adjustment preserves the education differences of interest and accounts for differences in age, sex, and birthplace within each skill group. Other differences in worker composition remain outside this check.

Migration Rates and Migrants Fact 4 uses IPUMS Census and ACS person records that report both current and prior-year MSA geography. The dependent variable equals one when the worker’s current MSA differs from her MSA one year earlier; the exact one-year samples are 1950, 2006, and 2008. Same-house respondents are zero, while unidentified, foreign, and invalid prior locations are excluded. The sample consists of employed individuals aged 25–55. Restricting the sample to 1950, 2006, and 2008 ensures that the dependent variable has the same one-year horizon and inter-MSA definition in every survey. Other available historical records report location five years earlier. A five-year residence-based mover indicator records whether the current location differs from the location five years earlier and therefore cannot be converted into a one-year binary outcome by dividing it by five. In a binary-response model, any positive rescaled value would still be classified as a move. I consequently exclude those observations rather than combine migration measures with different horizons. I pair the individual outcome with the contemporaneous skill ratio of the current MSA. The fullest specification includes state and survey-year fixed effects and covariates for sex, age group, and person weight; standard errors are clustered by current MSA.

Secular migration summarizes changes in the geographic distribution of each skill group. Its annual index is $R_t^k = 100 \sum_j |s_{jt}^k - s_{j,t-1}^k|$, where s_{jt}^k is city j ’s share of the national group- k population in calendar year t . This measure is distinct from the gross inter-MSA mover rate: offsetting flows can leave R_t^k near zero even when many individuals move. Gross movement is reported as a separate external fact and is not a calibration target.

A.2 Tables

Table A.1: Notation Used in the Empirical Analysis

Symbol	Definition
<i>Outcome variables</i>	
w_{kjt}	Log average wage of workers with skill k in region j at time t
$\Delta w_{kj(t_0, t_1)}$	Average annual log wage growth of skill- k workers in region j between years t_0 and t_1 (dependent variable in the β -convergence regressions)
SP_{jt}	Skill premium in region j at time t , $SP_{jt} \equiv w_{Hjt} - w_{Ljt}$
<i>Key regressors and controls</i>	
CollRatio_{jt}	Ratio of highly skilled to less skilled workers in region j at time t
$1(t = s)$	Indicator equal to one if observation year t corresponds to decade s ; used in equation 2
Pop_{jt}	Total population in region j at time t (log population used as a control in some regressions)
IndShare_{jdt}	Share of employment in industry d in region j at time t
RTI_{jvt}	Routine task intensity index for occupation v in region j at time t
x_j^{reg}	Index of regulatory constraints (e.g., housing or land-use regulation) in region j
x_j^{unav}	Measure of housing or land unavailability in region j
X_{jt}	Vector of additional regional controls used in some specifications
<i>Regression coefficients and fixed effects</i>	
β	Coefficient on initial log wages in the regional β -convergence regressions
β^H, β^L	β -coefficients estimated separately for highly skilled and less skilled wages in the convergence regressions
δ_s	Decade-specific coefficients on the interaction $\text{CollRatio}_{jt} \times 1(t = s)$ in the skill-premium regression, where $\text{CollRatio}_{jt} = H_{jt}/L_{jt}$ (equation 2)
α_j	Region fixed effect (when included)
γ_t	Time fixed effect (when included)

Note: Table A.1 summarizes the main empirical variables and regression parameters. All wages entering the regressions are expressed in logs. When symbols such as w_{kjt} or H_{jt} appear both here and in the structural-notation table below, they should be interpreted as the empirical counterparts of the corresponding model variables measured in the data.

Table A.2: Regional β -convergence Rates

Panel A: Convergence Rate						
	(1)	(2)	(3)	(4)	(5)	(6)
	1940–1980	1980–2010	1980–2010 IV	1940–1980	1980–2010	1980–2010 IV
$\ln(wage^{1940})$	−1.266*** (−12.44)			−1.589*** (−20.54)		
$\ln(wage^{1980})$		−0.122 (−0.25)	0.333 (0.46)		−0.891** (−2.91)	−0.164 (−0.25)
Panel B: Convergence Rate by Skills: β^H and β^L						
	(1)	(2)	(3)	(4)		
	No College	College	No College	College		
Panel B.1						
$\ln(wage^{1940})$	−1.235*** (−9.84)	−1.300*** (−6.42)				
$\ln(wage^{1980})$			−1.791*** (−6.48)	0.138 (0.38)		
Panel B.2						
$\ln(wage^{1940})$	−1.492*** (−16.66)	−2.093*** (−18.60)				
$\ln(wage^{1980})$			−2.128*** (−10.67)	−0.929*** (−3.18)		
N	127	127	239	239		

Note: Table A.2 reports the estimates of the β -convergence plotted in Figures 1 and 2. In Panel A, I report the estimate of the β coefficient for the whole sample underlying Figure 1. Column (1) reports the population-weighted β estimate for 1940–1980, and column (2) reports the corresponding estimate for 1980–2010. Columns (4) and (5) report the same specifications without population weights. Columns (3) and (6) report the population-weighted and unweighted IV estimates, respectively, using wages in 1970 as the instrument for initial wages in 1980. In Panel B.1, I report the estimates of the β -convergence corresponding to Figure 2. In column (1), I report the estimate for less skilled workers between 1940 and 1980; in column (2), for college graduates in the same period. In columns (3) and (4), the estimates are once again for the two groups, but for 1980–2010. Panel B.1 weights each skill-specific regression by the population of the corresponding skill group in the initial year of the window; Panel B.2 reports the same regressions without population weights. All standard errors are robust. The t -statistics are in parentheses. In Panel A, ***, **, and * denote statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. In both subpanels of Panel B, ***, **, *, and + denote significance at the 0.01, 0.05, 0.10, and 0.15 levels, respectively. The dependent variable in each regression is the annual average wage growth between the initial and final years reported at the top. All reported β -convergence coefficients are annual rates measured in percentage points.

Table A.3: Skill Premium by College Ratio of Cities over Time

	(1)	(2)
1(t=1940) x (H/L)	-0.55*** (-3.47)	-0.46*** (-3.30)
1(t=1950) x (H/L)	-0.22* (-2.04)	-0.12 (-1.46)
1(t=1960) x (H/L)	-0.24 (-1.40)	-0.14 (-0.99)
1(t=1970) x (H/L)	-0.07 (-0.60)	0.02 (0.18)
1(t=1980) x (H/L)	0.07 (1.29)	0.13 (1.80)
1(t=1990) x (H/L)	-0.12* (-2.31)	-0.05 (-0.77)
1(t=2000) x (H/L)	0.23** (2.91)	0.28** (2.95)
1(t=2010) x (H/L)	0.33*** (4.89)	0.38*** (4.49)
Population		3.64e-09* (2.48)

Note: Table A.3 reports the δ_s coefficient estimates for the historical specification in equation (2), using wages constructed from the IPUMS data. Before estimation, both the dependent variable (the skill premium, measured as the difference between the log wages of college graduates and less skilled workers) and the MSA college ratio are standardized within year, so the reported coefficients are standardized slopes. Both columns use population-weighted OLS with MSA and year fixed effects and standard errors clustered at the MSA level. Column (1) omits the raw population control; column (2) adds raw MSA population and is the specification plotted in Figure 3. The t -statistics are presented in parentheses. Asterisks ***, **, and * denote statistical significance at the 0.001, 0.01, and 0.05 levels, respectively.

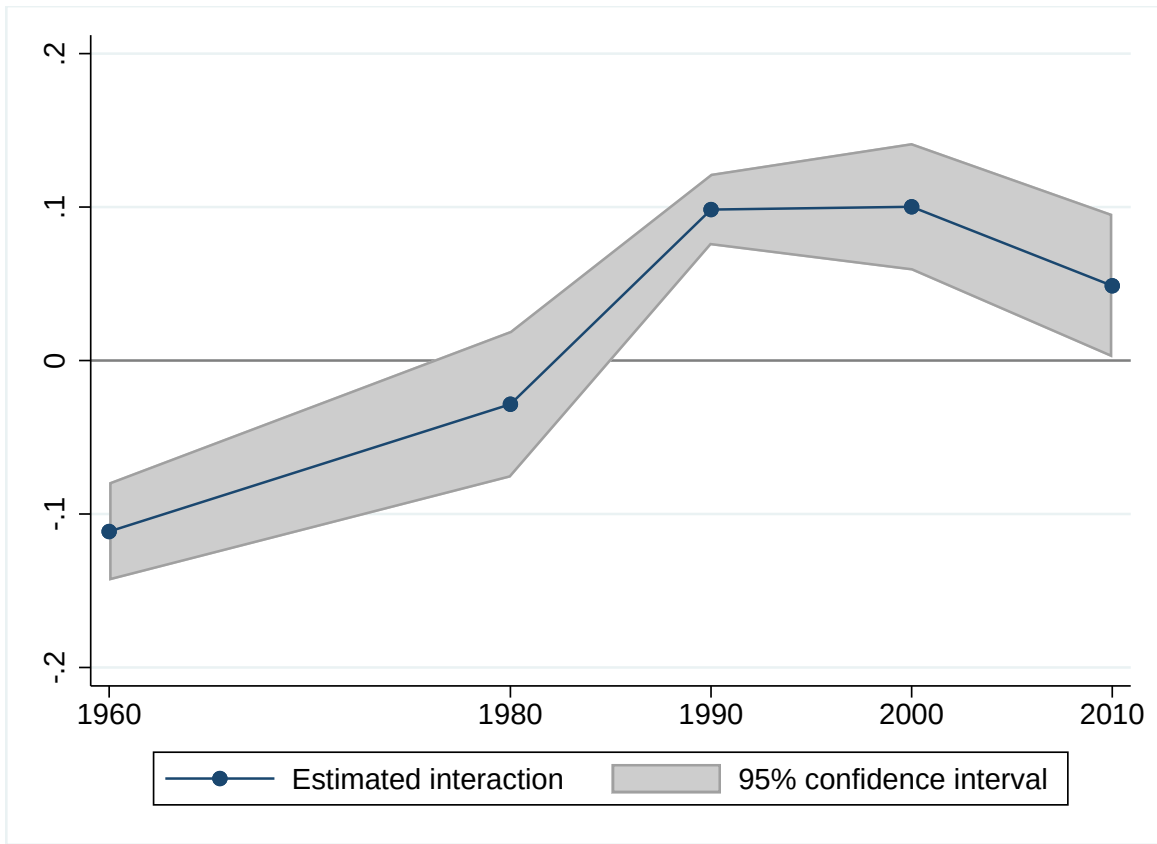
Table A.4: One-Year Inter-MSA Migration by College Ratio of Cities

<i>Exact one-year inter-MSA migration</i>				
	(1) A	(2) B	(3) C	(4) D
Inter-MSA mover				
1(t=1950) x 1(H) x (H/L)	0.105** (0.034)	-0.013 (0.028)	-0.020 (0.026)	-0.018 (0.025)
1(t=2006) x 1(H) x (H/L)	0.054** (0.019)	0.103*** (0.021)	0.081*** (0.018)	0.070*** (0.017)
1(t=2008) x 1(H) x (H/L)	0.092*** (0.021)	0.085*** (0.021)	0.068*** (0.019)	0.059** (0.018)
1(t=1950) x 1(H)	-0.233 (0.192)	0.469** (0.165)	0.514*** (0.153)	0.477** (0.148)
1(t=2006) x 1(H)	0.144*** (0.030)	0.057 (0.035)	0.096** (0.030)	0.127*** (0.028)
1(t=2008) x 1(H)	0.093** (0.034)	0.105** (0.034)	0.137*** (0.031)	0.158*** (0.030)
1(t=1950) x (H/L)	-9.300** (2.879)	5.355 (3.854)	2.848 (4.217)	2.880 (4.357)
1(t=2006) x (H/L)	-0.316* (0.132)	-0.438** (0.142)	-0.263** (0.086)	-0.281** (0.086)
1(t=2008) x (H/L)	-0.299* (0.121)	-0.283* (0.119)	-0.120 (0.063)	-0.137* (0.064)
Observations	1,454,321	1,454,321	1,454,321	1,454,321
Year FE	No	Yes	Yes	Yes
State FE	No	No	Yes	Yes
Controls	No	No	No	Yes

Note: Table A.4 reports equation (3) for exact one-year inter-MSA migration in 1950, 2006, and 2008; same-house respondents are zero. Parentheses are current-MSA-clustered standard errors; columns add year/state fixed effects and sex, age-group, and person-weight controls. Stars ***, **, and * denote 0.1, 1, and 5 percent significance.

Appendix Figure A.1 and Table A.5 provide a complementary March CPS comparison for 1960, 1980, 1990, 2000, and 2010. This longer sample uses the migration information available in each historical survey period and therefore does not maintain a common retrospective horizon across all decades. I do not combine these observations with the preferred one-year Census/ACS sample: one- and five-year binary mover indicators record different cumulative events, and dividing a five-year indicator by five does not recover a one-year migration outcome.

Figure A.1: Migration by Destination Education Level: March CPS



Note: The figure plots the estimated decade-specific interaction coefficient δ_s from column (4), the fullest March CPS specification of equation (3). This specification includes decade and state fixed effects and demographic controls, with standard errors clustered by MSA. The coefficients are reported for 1960, 1980, 1990, 2000, and 2010. The 1970 decade is omitted because the available CPS migration years around that window are sparse and discontinuous. Shaded regions are 95% confidence intervals; the horizontal line marks zero.

Table A.5: Migration by Destination Education Level: March CPS

	(1)	(2)	(3)	(4)
1(t=1960) x 1(H) x (H/L)	0.06 (1.53)	-0.09*** (-6.14)	-0.10*** (-7.06)	-0.11*** (-6.82)
1(t=1980) x 1(H) x (H/L)	-0.06 (-1.47)	-0.01 (-0.20)	-0.03 (-1.07)	-0.03 (-1.17)
1(t=1990) x 1(H) x (H/L)	0.10*** (5.87)	0.10*** (5.52)	0.10*** (8.35)	0.10*** (8.26)
1(t=2000) x 1(H) x (H/L)	0.14*** (5.95)	0.13*** (3.99)	0.12*** (5.61)	0.10*** (4.73)
1(t=2010) x 1(H) x (H/L)	0.27*** (10.17)	0.07* (2.39)	0.06* (2.37)	0.05* (2.04)
1(t=1960) x 1(H)	0.22 (1.42)	0.80*** (10.17)	0.84*** (11.76)	0.84*** (10.84)
1(t=1980) x 1(H)	0.42*** (4.74)	0.27*** (4.27)	0.33*** (5.86)	0.34*** (6.20)
1(t=1990) x 1(H)	0.04 (1.33)	0.03 (0.70)	0.04 (1.44)	0.05 (1.94)
1(t=2000) x 1(H)	-0.07 (-1.80)	-0.05 (-0.94)	-0.04 (-0.97)	-0.00 (-0.04)
1(t=2010) x 1(H)	-0.23*** (-4.55)	0.11* (2.03)	0.15** (3.04)	0.15** (3.13)
1(t=1960) x (H/L)	-1.38 (-1.17)	3.56*** (5.03)	2.00* (2.45)	1.93 (1.89)
1(t=1980) x (H/L)	1.37*** (5.20)	0.84* (2.06)	0.70* (2.13)	0.67* (2.08)
1(t=1990) x (H/L)	-0.30*** (-3.79)	-0.33*** (-5.34)	-0.33*** (-10.48)	-0.38*** (-11.58)
1(t=2000) x (H/L)	-0.22** (-3.18)	-0.19* (-2.28)	-0.20*** (-4.05)	-0.17*** (-3.50)
1(t=2010) x (H/L)	-0.30*** (-4.09)	0.07 (0.78)	0.03 (0.62)	0.06 (1.08)
Constant	-1.88*** (-49.62)	-2.35*** (-29.00)	-2.15*** (-22.39)	-1.41*** (-11.52)
Observations	1284338	1284338	1284338	1284338
Decade FE	No	Yes	Yes	Yes
State FE	No	No	Yes	Yes
Controls	No	No	No	Yes

Note: The table reports the complementary March CPS version of equation (3) using decade-specific interactions for 1960, 1980, 1990, 2000, and 2010. The 1970 decade is omitted because the available CPS migration years around that window are sparse and discontinuous. Columns progressively add decade fixed effects, state fixed effects, and demographic controls. Parentheses report MSA-clustered t -statistics. Stars ***, **, and * denote 0.1, 1, and 5 percent significance.

Table A.6: Annual Growth in $\log(H/L)$ vs. Initial $\log(H/L)$ in the Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1940–50	1950–60	1960–70	1970–80	1980–90	1990–2000	2000–10
Panel A							
$\frac{H}{L}$	-2.99***	-3.01***	-0.70**	0.84**	0.43**	0.97***	-0.41***
	(0.47)	(0.46)	(0.30)	(0.34)	(0.19)	(0.14)	(0.13)
	1940–60	1950–70	1960–80	1970–90	1980–2000	1990–2010	
Panel B							
$\frac{H}{L}$	-1.38***	-1.45***	0.31**	0.80***	0.70***		0.18 ⁺
	(0.14)	(0.26)	(0.14)	(0.23)	(0.12)		(0.11)
N	127	110	174	120	248		239

Note: Table A.6 reports population-weighted estimates of equation (4). The dependent variable is the annualized change in $\log(H/L)$, expressed in percentage points, and the regressor is the initial $\log(H/L)$. Panel A uses 10-year windows, and Panel B uses 20-year windows. Standard errors are in parentheses. The symbols +, *, **, and *** denote statistical significance at the 15, 10, 5, and 1 percent levels, respectively.

B Theory Appendix

Table B.1: Notation Used in the Structural Model

Symbol	Definition
<i>Indices</i>	
j	Region (MSA)
d	Industry/sector
$k \in \{H, L\}$	Skill group
t	Time period
i	Individual
<i>Population and Labor</i>	
H_{jt}, L_{jt}	Highly skilled and less skilled workers in region j
H_{djt}, L_{djt}	High- and low-skill labor in sector d in region j
N_{jt}	Population $H_{jt} + L_{jt}$
H_{vjt}, L_{vjt}	Employment in occupation v in region j
<i>Utility, Migration, Housing</i>	
T_{kjt}, O_{kjt}	Tradables and housing consumption of type k
u_{kjt}^C	Stone–Geary consumption index
$\nu_{kjt}(T, O)$	Deterministic preference index before consumption choice
$\tilde{u}_{ikjt}$	Flow utility in levels, $\exp(a_k \delta_{kjt}) \varepsilon_{ikjt}$
$\tilde{u}_{ikjt}(\bar{j}_t)$	Flow utility including inherited migration wedges
U_{ik}	Expected lifetime utility of individual i of skill k
A_{kjt}	Amenity for skill group k in region j
ζ_{ikjt}	Standard type-I extreme-value taste shock
ε_{ikjt}	Multiplicative taste shock $\exp(a_k \zeta_{ikjt})$
$m_k(j, j')$	Bilateral migration wedge in normalized choice units
$m_{k1}(j), m_{k2}(j')$	Origin and destination components of the normalized migration wedge
δ_{kjt}	Normalized indirect preference index
$M_k(j, j')$	Migration wedge in utility levels, $m_k(j, j')^{a_k}$
$V_{kt}(j)$	Remaining value in levels, with inherited multipliers removed
B_{kt}, Φ_{kt}	Aggregate choice index and origin-independent value factor
O_{jt}	Total housing demand/supply in region j
<i>Wages, Prices, and Costs</i>	
W_{Hjt}, W_{Ljt}	Skill-specific wages in j
R_{jt}	Rental price of housing
p_{djt}	Factory-gate price of intermediate good d in region j
$p_{j'jt}^d$	Delivered price in j' of good d from j
P_{djt}	Sectoral CES price index in j
P_{jt}	Final composite price index
c_{djt}	Unit cost of producing good d in region j
<i>Production and Trade</i>	
Y_{djt}	Output of intermediate good d in j
T_{jt}	Final tradable bundle
X_{jt}^W	Worker tradable expenditure net of subsistence housing in j
A_{jt}	Domestic absorption including locally owned housing income
E_{djt}	Expenditure on sector d in j
$X_{jj't}^d$	Shipment of good d from j to j'
$\tau_{jj'}^d$	Iceberg trade cost for sector d
<i>Technology and Productivity</i>	
η_{kdjt}	Efficiency units of skill- k labor in sector d
S_{Hjt}, S_{Ljt}	Local SBTC exposure shifters (for highly skilled and less skilled workers)

Continued on next page

Table B.1 – continued

Symbol	Definition
Λ_{kt}	Quantitative SBTC scaling term in the simulations
ξ_{kdjt}	Baseline log productivity
ε_{kdjt}	Fixed realized productivity innovation
<i>Structural Parameters</i>	
θ	Marginal expenditure share on tradables out of supernumerary income in Stone–Geary utility
$\bar{O}$	Subsistence housing level
β	Discount factor
$a_k \in (0, 1)$	Fixed scale relating the normalized index to utility levels
μ	Housing supply elasticity ($O_{jt} = R_{jt}^\mu$)
ρ	CES curvature between H and L
σ^{HL}	Elasticity of substitution $= 1/(1 - \rho)$
σ_d	Elasticity of substitution across varieties in sector d
α_d	Expenditure share of sector d in final-good aggregator
ω_d	Fixed share allocating location-level capital to sector d in the capital extension
μ_K	Labor share in value added in the capital extension
γ^H, γ^L	Productivity elasticity w.r.t. skill ratio H/L
ϕ^H, ϕ^L	Productivity elasticity w.r.t. city size N
λ, λ_t	Constant and time-varying SBTC shifts in the analytical appendix
$\lambda_k^{1960}, \lambda_k^{2010}$	Endpoint coefficients for alternative SBTC exposures
γ^p	Amenity elasticity w.r.t. local skill ratio
ϱ_H, ϱ_L	Skill-specific productivity persistence parameters
ψ	Intertemporal elasticity of substitution in the capital-accumulation extension

Note: Table B.1 summarizes the principal notation used in the structural model. Empirical-only notation (e.g., RTI, regulation indices, instruments, sample weights) is listed separately in the empirical appendix.

B.1 Definition of Equilibrium

Definition 3 (Sequential Competitive Equilibrium). Given:

- parameters $(\rho, \theta, \bar{O}, \mu, \{m_k(j, j'), \alpha_d, \sigma_d, \tau_{jj'}^d\}_{d,j,j'}, \gamma^H, \gamma^L, \phi^H, \phi^L, \varrho_H, \varrho_L, \gamma^p, \{a_k\}_k, \beta)$,
- initial worker locations consistent with the date-zero static equilibrium, with allocations $\{H_{j0}, L_{j0}\}_j$,
- an initial distribution of log productivities $\{\xi_{kdj0}\}_{k,d,j}$, fixed realized innovations $\{\varepsilon_{kdjt}\}_{k,d,j,t}$, and exogenous amenities $\{A_{kjt}\}_{k,j,t}$,
- national and local skill-biased technology inputs $\{S_{kjt}, \Lambda_{kt}\}_{k,j,t}$,

Here $a_k \in (0, 1)$ and $\beta \in (0, 1)$, and taste shocks are as defined in the main text. The time-invariant migration wedges enter utility levels as $M_k(j, j') = m_k(j, j')^{a_k}$. The quantitative implementation described in Subsection C.1 uses the fixed time-varying scaling term Λ_{kt} defined in the main text and imposes the restricted-L specification $\gamma^L = \phi^L = 0$. A *Sequential Competitive Equilibrium* is a sequence of allocations and prices

$$\left\{ W_{Hjt}, W_{Ljt}, R_{jt}, P_{djt}, P_{jt}, H_{jt}, L_{jt}, Y_{djt}, O_{jt}, \xi_{kdjt} \right\}_{j,d,t=0}^T$$

such that, for every location j , industry d , skill group $k \in \{H, L\}$, and time t :

1. **Household optimization.** Given wages, rents, prices and the aggregate path, consumption choices satisfy

$$T_{kjt} = \frac{\theta(W_{kjt} - R_{jt}\bar{O})}{P_{jt}}, \quad O_{kjt} = \theta\bar{O} + (1 - \theta)\frac{W_{kjt}}{R_{jt}},$$

with $W_{kjt} > R_{jt}\bar{O}$. Conditional on the given initial location, workers choose locations at dates $t = 1, \dots, T$ using current and past taste shocks. After removing inherited multipliers, the remaining problem for a worker entering date t in origin j is

$$V_{kt}(j) = \sup_{\{j_s\}_{s=t}^T} \mathbb{E} \left[\sum_{s=t}^T \beta^{s-t} \left\{ \prod_{\ell=t}^s M_k(j_{\ell-1}, j_\ell)^{-1} \right\} e^{a_k \delta_{kjss} \varepsilon_{ikjss}} \middle| j_{t-1} = j \right],$$

where the supremum is over these location policies. This problem satisfies the Bellman equation (9). Under time-invariant separable wedges, location probabilities and populations at each decision date are given by equations (15)–(16).

2. **Firm optimization.** In each sector d , firms hire H_{djt}, L_{djt} to minimize cost subject to the CES production function

$$Y_{djt} = [\eta_{Ldjt} L_{djt}^\rho + \eta_{Hdjt} H_{djt}^\rho]^{1/\rho},$$

taking wages and productivities η_{kdjt} as given. The resulting unit cost is

$$c_{djt}^{1-\sigma^{HL}} = \eta_{Ldjt}^{\sigma^{HL}} W_{Ljt}^{1-\sigma^{HL}} + \eta_{Hdjt}^{\sigma^{HL}} W_{Hjt}^{1-\sigma^{HL}}, \quad \sigma^{HL} = 1/(1 - \rho).$$

3. **Goods markets.** With iceberg trade costs $\tau_{jj'}^d$, sectoral Armington prices satisfy

$$P_{dj't} = \left[\sum_j (\tau_{jj'}^d p_{djt})^{1-\sigma_d} \right]^{\frac{1}{1-\sigma_d}}, \quad P_{jt} = \prod_d (P_{djt})^{\alpha_d},$$

where Armington origin weights are normalized to one in the quantitative implementation. The corresponding bilateral expenditure shares are

$$\pi_{jj't}^d = (\tau_{jj'}^d p_{djt})^{1-\sigma_d} (P_{dj't})^{\sigma_d-1}.$$

Under Stone–Geary preferences, worker tradable expenditure is

$$X_{jt}^W = \theta \left[(W_{Hjt} - R_{jt}\bar{O})H_{jt} + (W_{Ljt} - R_{jt}\bar{O})L_{jt} \right].$$

Local landlords spend housing income $R_{jt}^{\mu+1}$ on tradables, so domestic absorption is $\mathcal{A}_{jt} = X_{jt}^W + R_{jt}^{\mu+1}$. Implied sectoral expenditure is $E_{djt} = \alpha_d \mathcal{A}_{jt}$ and bilateral trade flows are $X_{jj't}^d = \pi_{jj't}^d E_{dj't}$. Under perfect competition and goods-market clearing, $p_{djt} = c_{djt}$ and $p_{djt}Y_{djt}$ equals domestic demand revenue plus ROW export revenue. When ROW is active, ROW expenditure is chosen so aggregate exports equal imports.

4. **Housing market.** Housing supply $O_{jt} = R_{jt}^\mu$ equals demand from all residents,

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

5. **Labor market clearing.** Aggregate employment of each skill equals the measure of residents: $H_{jt} = \sum_d H_{djt}$ and $L_{jt} = \sum_d L_{djt}$.

6. **Technology diffusion and realized innovations.** Log productivity evolves according to

$$\xi_{kdjt} = \varrho_k \xi_{kdj,t-1} + (1 - \varrho_k) \log \left(\frac{1}{N} \sum_{s=1}^N \exp(\xi_{kds,t-1}) \right) + \varepsilon_{kdjt},$$

where ϱ_H and ϱ_L are estimated separately and the realized innovations are fixed across counterfactuals.

7. **Trade-demand consistency and feasibility.** Equilibrium requires sectoral goods-market clearing, labor and housing clearing, the local budget identity including housing income, Armington share adding-up, and balanced aggregate ROW trade. All positive-support and Stone–Geary feasibility conditions must hold.

Remarks. A static equilibrium at any t satisfies all conditions holding t fixed. A sequential equilibrium is a path $\{\xi_{kdjt}, H_{jt}, L_{jt}, W_{kjt}, R_{jt}, P_{jt}\}_{t=0}^T$ consistent with household expectations, firms' optimality, market clearing, and the diffusion law of productivity. The next subsection separates static existence, conditional uniqueness, and numerical convergence; Appendix B.3 provides the full existence proof.

B.2 Existence and conditional uniqueness of equilibrium

This subsection gives the formal conditions and results summarized in Section 3.2. A static equilibrium at date t is defined in Section 3.1 and Appendix B.1. The dynamic model links successive static equilibria through the technology diffusion law (19).

Assumption 2. *Preferences are Stone–Geary with $\theta \in (0, 1)$, migration shocks are i.i.d. Type-I Extreme Value, and discounting satisfies $\beta \in (0, 1)$. Production functions are CES with $\sigma^{HL} > 1$ and $\sigma_d > 1$, trade costs $\tau_{jj'}^d \geq 1$ are finite, and housing supply is strictly increasing in rents with $\mu > 0$. Amenities A_{kjt} , log productivities ξ_{kdjt} , and SBTC shifters S_{kjt} are bounded above and below on the quantitative support.*

Step 1: Static existence.

Proposition 1 (Static existence). *Fix a date and consider the baseline without physical capital. Maintain positive finite fundamentals, positive national skill supplies, full domestic trade support at finite costs, common elasticities $\sigma > 1$ and $\mu > 0$, and $0 < \rho, \theta < 1$. Maintain the normalized location-choice equation and the aggregate-balanced reciprocal-Armington ROW rule, with the precise support and spending conditions stated in Appendix B.3. Suppose $\gamma^L = \phi^L = 0$, $\gamma^H, \phi^H, \gamma^p \geq 0$, and*

$$\gamma^H + \gamma^p \phi^H < 2 - \rho.$$

Then the stated static system, without an additional domestic price or income restriction, admits an equilibrium with both skills in every city and strictly positive supernumerary income. More precisely, such equilibria exist at every sufficiently large aggregate labor-income level.

Different aggregate labor-income levels yield distinct equilibria of the equations without an additional price or income restriction. The income index in the proof is a proof device, not a scale-neutral normalization.

Appendix B.3 constructs an explicit equilibrium residual in labor-income and population shares and proves its equivalence to the full clearing and optimality conditions. Its a priori estimates bound all shares away from zero. For sufficiently large aggregate labor income, a uniform housing bound makes the entire constructed domain strictly affordable. A regular reference equilibrium and the absence of boundary zeros then yield existence by invariance of Brouwer degree. This argument does not assume an invariant self-map or that populations alone uniquely determine wages and prices.

Step 2: Conditional uniqueness on the specified domain.

Assumption 3 (Conditional uniqueness criterion). *Fix a specified income or price restriction and a convex domain K_t of feasible independent equilibrium coordinates. Suppose equilibria satisfying that restriction and represented on K_t correspond one-to-one to fixed points of a map $\tilde{T}_t$ that is continuously differentiable on a neighborhood of K_t . Assume a uniform bound*

$$\sup_{x \in K_t} \|D\tilde{T}_t(x)\|_\infty = \sup_{x \in K_t} \max_i \sum_j \left| \frac{\partial \tilde{T}_{t,i}(x)}{\partial x_j} \right| \leq \bar{\kappa}_t < 1.$$

Proposition 2 (Conditional static uniqueness). *Under Assumption 3, there is at most one equilibrium satisfying the specified income or price restriction and represented on K_t .*

This argument maintains Assumption 3; it does not construct its map or domain from primitives. Fix the stated income or price restriction. For any $x, y \in K_t$, convexity places the segment joining them in K_t , and the integral mean-value formula gives

$$\tilde{T}_t(x) - \tilde{T}_t(y) = \int_0^1 D\tilde{T}_t(y + s(x - y))(x - y) ds.$$

The uniform row-sum bound therefore implies

$$\|\tilde{T}_t(x) - \tilde{T}_t(y)\|_\infty \leq \bar{\kappa}_t \|x - y\|_\infty.$$

If x and y are fixed points, $\bar{\kappa}_t < 1$ forces $x = y$. The assumed one-to-one equilibrium correspondence establishes Proposition 2. This is an at-most-one result on the specified domain, not an existence result under the quantitative price restriction. The spectral-radius bound

$$\rho(D\tilde{T}_t(x)) \leq \|D\tilde{T}_t(x)\|_\infty \leq \bar{\kappa}_t < 1$$

is a consequence of the uniform norm bound, not a substitute for it. The existence proof below uses an equilibrium residual and does not establish Assumption 3 for the solver's price restriction. In particular, it does not establish that small agglomeration elasticities alone bound the full equilibrium Jacobian.

Step 3: Dynamic path. Given initial log productivity, the fixed realized innovations, and the exogenous shifters, equation (19) determines log productivities by forward iteration. Applying Proposition 1 at each date for which its assumptions hold gives a sequence of static equilibria without an additional domestic price or income restriction. A unique sequence under a specified restriction would require existence and a separate static-uniqueness result

under that restriction at every date. The realized innovations are recovered as residuals of the productivity transition, with no separate structural identification.

Step 4: Numerical verification. The quantitative implementation fixes the producer price in city 1, sector 1 to one, $p_{1,1,t} = 1$. Proposition 1 does not certify existence or uniqueness under this additional restriction. Neither the fixed-point representation in Assumption 3 nor its uniform contraction bound is established here for this price restriction. Solver convergence and equilibrium residual checks provide numerical evidence for the computed paths.

The previously reported numerical convergence checks are separate from the existence proof and the conditional uniqueness criterion. Section C.1 describes the fixed-point algorithm. At the estimated parameter values, the algorithm converges from the tested starting values. When the agglomeration parameters are increased beyond the estimated range, it can fail to converge or generate allocations with extreme concentration of activity. These checks document the behavior of the numerical routine; they do not constitute a global uniqueness proof or establish the conditional uniqueness assumptions for the quantitative price restriction.

B.3 Proof of static existence

Fix a date in the system of Appendix B.1. There are finite $J, D \geq 1$, national skill totals $\bar{H}, \bar{L} > 0$, and strictly positive finite coefficients. Write $n_{Hj} = H_j$, $n_{Lj} = L_j$, $N_j = H_j + L_j$, $r_j = H_j/L_j$, and $\bar{N} = \bar{H} + \bar{L}$. Assume $\gamma^L = \phi^L = 0$, $\gamma^H, \phi^H, \gamma^p, b \geq 0$, $b = \bar{O}$, $0 < \rho, \theta < 1$, and common $\sigma > 1$, $\mu > 0$. Sector weights are fixed, positive, and sum to one. Domestic trade has full support and unit Armington origin weights. ROW prices are fixed and positive; every ROW-open sector connects each city to ROW by finite import and export costs. Armington denominators sum over all domestic origins and also ROW exactly in open sectors, with fixed p_{d0} . Both skills share θ, γ^p . With $s = (1 - \rho)^{-1}$, use the CES unit cost p_{dj} of B.1 and factor shares $f_{kdj} = \eta_{kdj}^s w_{kj}^{1-s} p_{dj}^{s-1}$, technology $\eta_{Hdj} = a_{Hdj} r_j^{\gamma^H} N_j^{\phi^H}$, $\eta_{Ldj} = a_{Ldj}$, and normalized location index $V_{kj} = c_{kj} r_j^{\gamma^p} (w_{kj} - bR_j) / (P_j^\theta R_j^{1-\theta})$, with $n_{kj} = \bar{n}_k V_{kj} / \sum_i V_{ki}$ and $\bar{n}_H = \bar{H}$, $\bar{n}_L = \bar{L}$.

For income $E_j = \sum_k w_{kj} n_{kj}$, housing clearing is $R_j^{\mu+1} = \theta b R_j N_j + (1 - \theta) E_j$; hence worker plus landlord tradable spending equals E_j . For a nonempty ROW-open set $\mathcal{A}$, let $I = \sum_{\ell, d \in \mathcal{A}} \alpha_d \pi_{0\ell}^d E_\ell$ and $q_j^d = (\tau_{j0}^d p_{dj})^{1-\sigma} / \sum_i (\tau_{i0}^d p_{di})^{1-\sigma}$. Any prescribed continuous balanced reciprocal Armington ROW rule is represented by continuous $\omega_d \geq 0$, $\sum_{d \in \mathcal{A}} \omega_d = 1$: sectoral revenue is $D_{dj} = \sum_\ell \alpha_d \pi_{j\ell}^d E_\ell + I \omega_d q_j^d$. Export terms vanish in closed sectors; with no ROW, omit them and I . Thus $\sum_{dj} D_{dj} = \sum_j E_j$ at every positive candidate. Define

$$\Delta = 2 - \rho - \gamma^H, \quad u = \phi^H / \Delta, \quad t = (\sigma - 1) / \sigma, \quad q = (\mu + \theta) / (\mu + 1), \quad \mathcal{B} = 2 - \gamma^p u - tq.$$

Proposition 3 (Existence result). *Under the conditions above, suppose $\Delta > 0$ and $\mathcal{B} > 0$. There is a finite primitive-dependent M_* such that, for every $M > M_*$, the static system admits an equilibrium with $\sum_j E_j = M$, $H_j, L_j > 0$, and $w_{kj} - bR_j \geq \epsilon_M > 0$ for every k, j .*

All domestic sectoral outputs and both sectoral skill inputs are positive. If $b = 0$, the result holds for every $M > 0$.

Corollary 1 (A simpler sufficient restriction). *The proposition applies if $\gamma^H + \gamma^p \phi^H < 2 - \rho$, for every common $\sigma > 1$ and $\mu > 0$.*

Indeed, this restriction gives $\Delta > 0$, $\gamma^p u < 1$, and $\mathcal{B} > 0$ since $0 < t, q < 1$. The proof fixes M without omitting any clearing equation or assuming allocation invariance. It does not cover an arbitrary additional income or price anchor, or assert target uniqueness or contraction.

1. Equilibrium coordinates and the reference path. For positive shares $\sum_{kj} v_{kj} = 1$, $\sum_j z_{kj} = 1$, reconstruct $n_{kj} = \bar{n}_k z_{kj}$, $w_{kj} = M v_{kj} / n_{kj}$, and $E_j = M(v_{Hj} + v_{Lj})$. The housing root is unique and smooth: $R^\mu - \theta b N_j - (1 - \theta) E_j / R$ increases from $-\infty$ to $+\infty$ with positive derivative. CES and Armington formulas and the ROW rule determine the remaining objects. On the affordable domain define $T_{kj}^v = \sum_d f_{kdj} D_{dj} / M$ and $T_{kj}^z = V_{kj} / \sum_i V_{ki}$, continuous positive probability vectors by the budget identity. For $\ell(x) = (\log(x_i / x_m))_{i < m}$, set $y = (\ell(v), \ell(z_H), \ell(z_L)) \in \mathbb{R}^{4J-3}$. The income coordinates $\Phi(v) = ((\log(v_{Hj} / v_{Lj}))_j, (\log[(v_{Hj} + v_{Lj}) / (v_{HJ} + v_{LJ})])_{j < J})$ smoothly and bijectively identify the interior income simplex. Use the square residual

$$F_M(y) = (\Phi(v) - \Phi(T^v), \ell(z_H) - \ell(T_H^z), \ell(z_L) - \ell(T_L^z)).$$

Its zeros give $v = T^v$, $z = T^z$. Set $Y_{dj} = D_{dj} / p_{dj}$, $n_{kdj} = f_{kdj} D_{dj} / w_{kj}$: CES cost minimization gives these outputs and zero profits, and $\sum_d n_{kdj} = M v_{kj} / w_{kj} = n_{kj}$ clears factors. Location shares and Stone–Geary demands are optimal and strictly feasible; housing clears by construction. Delivered demand times iceberg costs sums to Y_{dj} , clearing goods; absorption and the ROW rule complete budgets. Conversely, equilibrium cost minimization, factor and location clearing give these fixed-point equations. All objects are reconstructed uniquely from shares, establishing exact correspondence.

Connect the reference economy to the model by multiplying $(\gamma^H, \phi^H, \gamma^p, b)$ by $\lambda \in [0, 1]$ and replacing positive a_{kdj}, c_{kj}, τ by their λ th powers. Keep all other primitives and the prescribed continuous ROW rule fixed. Then $\Delta_\lambda := 2 - \rho - \lambda \gamma^H \geq \Delta > 0$, $0 \leq u_\lambda := \lambda \phi^H / \Delta_\lambda \leq u$, and $\mathcal{B}_\lambda = 2 - \lambda \gamma^p u_\lambda - tq \geq \mathcal{B} > 0$. Below $X \asymp Y$ means two-sided positive bounds depending only on primitives and these positive margins, uniformly in M, λ and equilibrium.

2. Bounds on equilibrium shares. Initially consider any equilibrium satisfying $\lambda b R_j / w_{kj} \leq 1/2$. Factor clearing gives $w_{kj}^s n_{kj} = \sum_d \eta_{kdj}^s p_{dj}^{s-1} D_{dj}$, hence $w_{Hj} / w_{Lj} \asymp r_j^{\lambda \gamma^H + \rho - 1} N_j^{\lambda \phi^H}$. Location ratios give $r_j = C \nu_j w_{Hj} / w_{Lj}$, where $C = (\bar{H} / \bar{L}) \sum_i V_{Li} / \sum_i V_{Hi}$ and $\nu_j \asymp 1$. Therefore $r_j = K b_j N_j^{u_\lambda}$ with $K = C^{1/\Delta_\lambda}$ and $B^{-1} \leq b_j \leq B$ for a primitive-only $B \geq 1$. To bound K without a population floor, put $U = \max(1, \bar{N}^u)$, $e = \bar{L} / (2J)$, $L_0 = \min(1, e^u)$. Then $\bar{H} = \sum_j N_j r_j / (1 + r_j) \leq \bar{N} K B U$. Cities with $N_j < e$ contain less than half the low-skilled workers; on the others $N_j^{u_\lambda} \geq L_0$, so $\bar{L} / 2 \leq \sum_{N_j \geq e} N_j / (1 + r_j) \leq \bar{N} / (1 + K B^{-1} L_0)$. These inequalities bound K above and away from zero. Consequently $r_j \asymp N_j^{u_\lambda}$, r_j is bounded above, $w_{Hj} / w_{Lj} \asymp r_j$, and $E_j \asymp w_{Lj} N_j$.

Put $A_j = r_j^{\lambda \gamma^H} N_j^{\lambda \phi^H}$, $Q_j = (A_j H_j^\rho + L_j^\rho)^{1/\rho}$ and $p_j^0 = (A_j^s w_{Hj}^{1-s} + w_{Lj}^{1-s})^{1/(1-s)}$. The coefficients a_{kdj}^λ are bounded above and away from zero; for $F_j(h, l) = (A_j h^\rho + l^\rho)^{1/\rho}$, concavity and $0 < \rho < 1$ give $\sum_d F_j \leq Q_j$ and $Q_j^0 \leq \sum_d F_j^0 \leq D^{1-\rho} (\sum_d F_j)^\rho$. Thus $E_j \asymp p_j^0 Q_j$. Also $Q_j = N_j [1 + r_j^{\lambda \gamma^H + \rho} N_j^{\lambda \phi^H}]^{1/\rho} / (1 + r_j) \asymp N_j$, using $\lambda \phi^H + u_\lambda (\lambda \gamma^H + \rho) = 2u_\lambda$ and the upper bounds just proved. For every domestic destination and sector, $\pi_{j\ell}^d / \pi_{i\ell}^d = (\tau_{j\ell}^d p_{dj} / \tau_{i\ell}^d p_{di})^{1-\sigma}$; the same ratio identity holds for conditional exports q_j^d / q_i^d . Uniformly bounded trade and sectoral cost factors, multiplied by common nonnegative expenditures and summed, give $E_j / E_i \asymp (p_j^0 / p_i^0)^{1-\sigma}$ independently of ROW allocations, including zero export weights. Eliminating prices with $E_j \asymp p_j^0 Q_j$ yields $E_j / E_i \asymp (N_j / N_i)^t$.

Housing gives $(1 - \theta)E_j \leq R_j^{\mu+1} \leq (1 - \theta/2)E_j$. Delivered-price ratios across destinations are uniformly bounded, including ROW; hence $P_j/P_i \asymp 1$ and $R_j/R_i \asymp (E_j/E_i)^{1/(\mu+1)}$. Location then gives $V_{Lj}/V_{Li} \asymp (N_j/N_i)^{\lambda\gamma^p u_\lambda - 1 + tq}$. Since $V_{Lj}/V_{Li} = L_j/L_i \asymp N_j/N_i$, $(N_j/N_i)^{\mathcal{B}\lambda} \asymp 1$. The positive margin $\mathcal{B}$ and fixed national totals bound all N_j away from zero; the bounds on r_j then bound H_j, L_j away from zero. Income ratios bound E_j/M away from zero, and $w_{Hj}/w_{Lj} \asymp r_j$ bounds both within-city skill-income fractions away from zero. Thus all $v_{kj}, z_{kj} \geq \epsilon_0 > 0$, with $\epsilon_0 \leq 1/2$ depending only on primitives, and $\|y\|_\infty \leq C_0 := \log(1/\epsilon_0)$.

3. A feasible domain. Set $C = C_0 + 1$, $\Omega = (-C, C)^{4J-3}$, and $\epsilon_v = e^{-C}/[1 + (2J - 1)e^C]$. Inverting the log ratios gives $v_{kj} \geq \epsilon_v$, hence $w_{kj} \geq M\epsilon_v/\bar{N}$ on $\bar{\Omega}$. At every candidate and every λ , $E_j \leq M$, $N_j \leq \bar{N}$, and housing implies

$$R_j \leq R_{\max}(M) := \max\{(2\theta b\bar{N})^{1/\mu}, [2(1 - \theta)M]^{1/(\mu+1)}\}.$$

A rent exceeding both terms contradicts housing clearing. Choose finite M_* so $b\bar{N}R_{\max}(M)/(M\epsilon_v) < 1/2$ for all $M > M_*$; this is possible because $R_{\max}(M)/M \rightarrow 0$. For $b = 0$ take $M_* = 0$. Every candidate then satisfies $\lambda b R_j/w_{kj} < 1/2$ and $w_{kj} - \lambda b R_j > M\epsilon_v/(2\bar{N}) > 0$. Thus $F_{M,\lambda}$ is continuous on $\bar{\Omega} \times [0, 1]$; every zero reconstructs an equilibrium and has norm at most $C_0 < C$. No boundary zero occurs.

4. The reference zero and continuation. At $\lambda = 0$, technologies and active trade costs are one and $b_0 = 0$. Sectors share the city cost p_j and factor shares. Each domestic and conditional export revenue term is proportional to $p_j^{1-\sigma}$ with a coefficient common across origins. The budget identity therefore gives $\sum_d D_{dj} = M p_j^{1-\sigma} / \sum_i p_i^{1-\sigma}$ at every positive candidate, for any continuous ROW allocation rule and fixed foreign prices. The residual blocks are exactly

$$\begin{aligned} G_j^s &= s \log(w_{Hj}/w_{Lj}) + \log(H_j/L_j), \\ G_j^E &= \log(E_j/E_J) - (1 - \sigma) \log(p_j/p_J), \quad j < J, \\ G_j^k &= \log(n_{kj}/n_{kJ}) - \log(w_{kj}/w_{kJ}) + \frac{1-\theta}{\mu+1} \log(E_j/E_J), \quad j < J. \end{aligned}$$

Thus the reference residual is smooth even for merely continuous ROW weights. A zero has $H_j = h := \bar{H}/J$, $L_j = l := \bar{L}/J$, $E_j = M/J$, $v_{Hj} = \omega/J$, $v_{Lj} = (1 - \omega)/J$, $z_{kj} = 1/J$, where $\omega = h^\rho/(h^\rho + l^\rho)$; reconstruction verifies clearing and affordability. At any zero, skill and location equations imply $w_{Hj}/w_{Lj} = r_j^{\rho-1}$ and $r_j^{2-\rho} = C$, so national totals fix $r_j = \bar{H}/\bar{L}$. Identical factor proportions give $Q_j = (H_j^\rho + L_j^\rho)^{1/\rho} \propto N_j$, $E_j = p_j Q_j$, and $E_j/E_i = (N_j/N_i)^t$. Wages scale as E_j/N_j , rents as $E_j^{1/(\mu+1)}$, and destination price indices coincide. Location gives $(N_j/N_i)^{2-tq} = 1$. Since $2 - tq > 0$, the zero is unique and lies inside Ω .

For a Jacobian-kernel vector, write logarithmic differentials $h_j = d \log H_j$, $l_j = d \log L_j$, $e_j = d \log E_j$, $u_j = d \log(w_{Hj}/w_{Lj})$, and $\zeta_j = \omega h_j + (1 - \omega)l_j$. Fixed totals imply zero sums of h, l, e, ζ . The income identity gives $d \log w_{Hj} = e_j - \zeta_j + (1 - \omega)u_j$, $d \log w_{Lj} = e_j - \zeta_j - \omega u_j$. Differentiating skill equations gives $u_j + (1 - \rho)(h_j - l_j) = 0$, hence $\sum_j u_j = 0$. Unit costs give $d \log p_j = e_j - \zeta_j$; the income residual and zero means yield $e_j = t\zeta_j$. Population residuals give $h_j + \zeta_j - qe_j - (1 - \omega)u_j = 0$ and $l_j + \zeta_j - qe_j + \omega u_j = 0$. Their difference and weighted mean yield $h_j - l_j - u_j = 0$ and $2\zeta_j - qe_j = 0$. Thus $(2 - \rho)(h_j - l_j) = 0$, $u_j = 0$, and $(2 - tq)\zeta_j = 0$: every differential vanishes and the square Jacobian is nonsingular. For $J = 1$, fixed totals set $h_1 = l_1 = e_1 = 0$ and the skill equation sets $u_1 = 0$.

For $M > M_*$, the regular reference zero has degree $\text{sign det } D_y F_{M,0} \in \{-1, 1\}$. Continuity and absence of boundary zeros give $\deg(F_{M,1}, \Omega, 0) = \deg(F_{M,0}, \Omega, 0) \neq 0$, hence a target zero (Benevieri et al., 2023, Definition 3.3 and Theorems 3.9, 3.11). Exact correspondence supplies the full equilibrium and the uniform positive affordability margin. $\square$

C Additional Details on Model Parametrization

This section provides computational, measurement, and diagnostic details that complement the model in Section 3 and the quantitative exercises in the main text. I address these details in the following parts:

In Subsection C.1, I describe the numerical procedure used to compute the simulation sequence of spatial equilibria from 1960 to 2010 under the equilibrium conditions above. In Subsection C.2, I derive the static multi-industry labor-demand aggregation and distinguish it from the sector-level productivity transition. Subsection D.4 documents the five SBTC constructions. Subsection C.4 compares the recovered amenities and destination components of bilateral migration costs with observed city characteristics. Subsection C.5 discusses an external migration-cost benchmark. Subsection C.6 documents the full productivity, centered-wage, and amenity inversion; Table C.9 summarizes only its 1960 productivity level. Finally, Subsection C.7 reports the corresponding spatial moments in the model and data.

C.1 Description of the Computational Algorithm

This subsection describes the numerical procedure used to solve the model period by period and generate the full transition path from 1960 to 2010. The procedure solves a sequence of static spatial equilibria with endogenous migration, Armington trade across U.S. locations, and an optional Rest of World (ROW) node that opens in 1980. For transparency, I briefly restate below the subset of equilibrium equations on which the solver iterates. They correspond to the main-text objects in (16), (6), (17), (18), (23), and (25); the numerical price restriction is stated below.

At each date t , the unknowns are $\{W_{Hjt}, W_{Ljt}, R_{jt}, P_{jt}, P_{djt}, p_{djt}, H_{djt}, L_{djt}, H_{jt}, L_{jt}\}_{j=1, d=1}^{N, D}$. The sectoral production block is

$$Y_{djt} = \left(\eta_{Ldjt} L_{djt}^\rho + \eta_{Hdjt} H_{djt}^\rho \right)^{1/\rho}, \text{ with}$$

$$\eta_{Hdjt} = \left(\frac{H_{jt}}{L_{jt}} \right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H} S_{Hjt}^{\Lambda_{Ht}} \exp(\xi_{Hdjt} + b_{Hjt}), \quad \eta_{Ldjt} = S_{Ljt}^{\Lambda_{Lt}} \exp(\xi_{Ldjt} + b_{Ljt}),$$

where the second expression imposes $\gamma^L = \phi^L = 0$. The log productivities (ξ_{Hdjt}, ξ_{Ldjt}) follow equation (19); realized innovations and productivity residuals are fixed across counterfactuals. The SBTC block determines the path of $(S_{Hjt}, S_{Ljt}, \Lambda_{Ht}, \Lambda_{Lt})$. Trade is solved industry by industry:

$$P_{dj't}^{1-\sigma_d} = \sum_{j=1}^N (\tau_{jj'}^d p_{djt})^{1-\sigma_d} + \mathbf{1}\{t \geq t^{ROW}\} (\tau_{ROW,j'}^d p_d^{ROW})^{1-\sigma_d},$$

and the aggregate non-housing price index is

$$\log P_{j't} = \sum_{d=1}^D \alpha_d \log P_{dj't}.$$

In the code, domestic bilateral trade costs are distance-based iceberg costs using the cleaned distance matrix:

$$\tau_{jj'}^d = \min\{\max(\text{dist}_{jj'}^\kappa, 1), \bar{\tau}\}, \quad \kappa = 0.5, \bar{\tau} = 50.$$

In the fixed-point routine, sectoral labor demand and wage clearing are written in terms of producer prices p_{djt} ; destination price indices P_{djt} enter through Armington demand and indirect utility. Worker tradable expenditure is

$$X_{jt}^W = \theta \left[(W_{Hjt} - R_{jt}\bar{O})H_{jt} + (W_{Ljt} - R_{jt}\bar{O})L_{jt} \right].$$

Total domestic absorption is $\mathcal{A}_{jt} = X_{jt}^W + R_{jt}^{\mu+1}$. Producer prices equal unit costs and sectoral goods clearing requires

$$p_{djt}Y_{djt} = \sum_{j'=1}^N \pi_{jj't}^d \alpha_d \mathcal{A}_{j't} + X_{jt}^{d,ROW}.$$

When ROW is active, its spending is determined endogenously so aggregate export revenue equals imports. Local housing income is part of domestic absorption; in the capital extension, locally owned capital income is added as well.

The numerical routine fixes the producer price in city 1, sector 1 to one, $p_{1,1,t} = 1$, and expresses the supplied ROW prices in these units. This is an additional price restriction. Proposition 1 proves existence without that restriction and does not certify existence or uniqueness at the selected price. The reported convergence and residual checks are numerical diagnostics for the computed solutions.

Fixed-point routine (within each year t).

1. Initialize from previous-year values for $(H, L, H_d, L_d, W_H, W_L, R, p, P_d, P)$.
2. Carry (ξ_H, ξ_L) forward using the skill-specific persistence and realized innovations, then combine those log productivities with the selected SBTC and agglomeration blocks and the fixed productivity residuals.
3. Given current guesses, compute unit costs, Armington shares, domestic and ROW demand, goods-clearing wages, rents, utility, and migration shares holding the amenity vector fixed.

4. Apply damped positive updates to wages and allocations, preserving the national skill totals used in the model.
5. Iterate until goods, production, household-budget, housing, factor-clearing, allocation, national-total, and ROW-balance residuals satisfy the convergence criteria.

The same goods-market-clearing conditions are used across non-capital specifications. Amenities, productivity residuals, realized innovations, and all parameters other than those changed by the counterfactual are held fixed. Agglomeration shutdowns preserve the 1980 city-size stock and phase out only the post-1980 ϕ^H flow according to the paper schedule; the pre-1980 path is unchanged.

C.2 Aggregation of the Static Labor-Demand Block

This subsection shows how the static multi-industry labor-demand system can be aggregated exactly to the city $\times$ time wage equations used in the GMM. The key implication is that separate observations on sectoral prices or sectoral outputs are not needed in the empirical wage equations, because those objects enter only through city-level composite residuals.

Lemma 1 (Exact aggregation of the static labor-demand block). *Consider a location j at time t with industries $d = 1, \dots, D$. Suppose:*

1. *Common city-level wages across industries. High- and low-skill labor are mobile across industries within location j , so equilibrium wages are common across sectors:*

$$W_{Hjt} = W_{Hdj t} \quad \text{and} \quad W_{Ljt} = W_{Ldj t} \quad \forall d.$$

2. *Common propagation coefficients across industries. The coefficients governing the response of productivity to local skill composition and local scale do not vary by industry:*

$$\gamma_d^H = \gamma^H, \quad \gamma_d^L = \gamma^L, \quad \phi_d^H = \phi^H, \quad \phi_d^L = \phi^L,$$

and the SBTC scaling terms Λ_{Ht} and Λ_{Lt} are common across industries within a city and date.

3. *Multiplicative sectoral productivity decomposition. Sectoral productivity can be written as*

$$\eta_{Hdj t} = A_{Hjt} \exp(\xi_{Hdj t} + b_{Hjt}), \quad \eta_{Ldj t} = A_{Ljt} \exp(\xi_{Ldj t} + b_{Ljt}),$$

where b_{kjt} is common across industries and

$$A_{Hjt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H} S_{Hjt}^{\Lambda_{Ht}}, \quad A_{Ljt} = \left(\frac{H_{jt}}{L_{jt}}\right)^{\gamma^L} (H_{jt} + L_{jt})^{\phi^L} S_{Ljt}^{\Lambda_{Lt}},$$

so that the city-level components are common across industries.

4. CES production and perfect competition. In each industry,

$$Y_{djt} = \left(\eta_{Ldjt} L_{djt}^\rho + \eta_{Hdjt} H_{djt}^\rho \right)^{1/\rho}, \quad \rho < 1,$$

and firms are perfectly competitive, so wages equal marginal products.

5. Interior solutions and positive prices/outputs. For all active industries, sectoral prices and outputs satisfy

$$p_{djt} > 0, \quad Y_{djt} > 0,$$

and the sectoral labor allocations satisfy

$$H_{djt} > 0, \quad L_{djt} > 0.$$

Let

$$\sigma^{HL} \equiv \frac{1}{1 - \rho}.$$

Then the sectoral first-order conditions aggregate exactly to the city-level wage equations

$$\ln W_{Hjt} = \gamma^H \ln \left(\frac{H_{jt}}{L_{jt}} \right) + \phi^H \ln(H_{jt} + L_{jt}) + \Lambda_{Ht} \ln S_{Hjt} + (\rho - 1) \ln H_{jt} + \Xi_{Hjt}, \quad (34)$$

$$\ln W_{Ljt} = \gamma^L \ln \left(\frac{H_{jt}}{L_{jt}} \right) + \phi^L \ln(H_{jt} + L_{jt}) + \Lambda_{Lt} \ln S_{Ljt} + (\rho - 1) \ln L_{jt} + \Xi_{Ljt}, \quad (35)$$

where the city-level composite residuals are

$$\Xi_{Hjt} \equiv b_{Hjt} + \frac{1}{\sigma^{HL}} \ln \left[\sum_{d=1}^D Y_{djt} \left(p_{djt} \exp(\xi_{Hdjt}) \right)^{\sigma^{HL}} \right], \quad (36)$$

$$\Xi_{Ljt} \equiv b_{Ljt} + \frac{1}{\sigma^{HL}} \ln \left[\sum_{d=1}^D Y_{djt} \left(p_{djt} \exp(\xi_{Ldjt}) \right)^{\sigma^{HL}} \right]. \quad (37)$$

Hence, under assumptions 1–5, the static multi-industry labor-demand system can be represented exactly at the city×time level through the composite residuals Ξ_{Hjt} and Ξ_{Ljt} .

Proof. Under perfect competition, the sectoral first-order conditions for each industry d are

$$W_{Hjt} = p_{djt} \eta_{Hdj} Y_{djt}^{1-\rho} H_{djt}^{\rho-1}, \quad W_{Ljt} = p_{djt} \eta_{Ldj} Y_{djt}^{1-\rho} L_{djt}^{\rho-1}.$$

Rearranging yields sectoral labor-demand schedules

$$H_{djt} = Y_{djt} \left(\frac{p_{djt} \eta_{Hdj}}{W_{Hjt}} \right)^{\sigma^{HL}}, \quad L_{djt} = Y_{djt} \left(\frac{p_{djt} \eta_{Ldj}}{W_{Ljt}} \right)^{\sigma^{HL}},$$

where $\sigma^{HL} = 1/(1 - \rho)$.

Summing across industries and using the fact that wages are common across industries within city j gives

$$H_{jt} = W_{Hjt}^{-\sigma^{HL}} \sum_{d=1}^D Y_{djt} (p_{djt} \eta_{Hdj})^{\sigma^{HL}},$$

$$L_{jt} = W_{Ljt}^{-\sigma^{HL}} \sum_{d=1}^D Y_{djt} (p_{djt} \eta_{Ldj})^{\sigma^{HL}}.$$

Solving for wages,

$$W_{Hjt} = \left(\frac{\sum_{d=1}^D Y_{djt} (p_{djt} \eta_{Hdj})^{\sigma^{HL}}}{H_{jt}} \right)^{1/\sigma^{HL}}, \quad W_{Ljt} = \left(\frac{\sum_{d=1}^D Y_{djt} (p_{djt} \eta_{Ldj})^{\sigma^{HL}}}{L_{jt}} \right)^{1/\sigma^{HL}}.$$

Now substitute

$$\eta_{Hdj} = A_{Hjt} \exp(\xi_{Hdj} + b_{Hjt}), \quad \eta_{Ldj} = A_{Ljt} \exp(\xi_{Ldj} + b_{Ljt}),$$

with A_{Hjt} , A_{Ljt} , b_{Hjt} , and b_{Ljt} common across industries. Factoring out the city-level terms gives:

$$W_{Hjt} = A_{Hjt} \exp(b_{Hjt}) \left(\frac{\sum_{d=1}^D Y_{djt} (p_{djt} \exp(\xi_{Hdj}))^{\sigma^{HL}}}{H_{jt}} \right)^{1/\sigma^{HL}},$$

$$W_{Ljt} = A_{Ljt} \exp(b_{Ljt}) \left(\frac{\sum_{d=1}^D Y_{djt} (p_{djt} \exp(\xi_{Ldj}))^{\sigma^{HL}}}{L_{jt}} \right)^{1/\sigma^{HL}}.$$

Taking logs and using the definitions of A_{Hjt} and A_{Ljt} yields

$$\ln W_{Hjt} = \gamma^H \ln \left(\frac{H_{jt}}{L_{jt}} \right) + \phi^H \ln(H_{jt} + L_{jt}) + \Lambda_{Ht} \ln S_{Hjt}$$

$$- \frac{1}{\sigma^{HL}} \ln H_{jt} + b_{Hjt} + \frac{1}{\sigma^{HL}} \ln \left[\sum_{d=1}^D Y_{djt} (p_{djt} \exp(\xi_{Hdj}))^{\sigma^{HL}} \right].$$

The low-skill wage follows analogously. Since $1/\sigma^{HL} = 1 - \rho$, we have $-\frac{1}{\sigma^{HL}} = \rho - 1$, which gives (34)–(35) with (36)–(37). $\square$

This result motivates the city-level residual representation used in the GMM, but the estimation itself is implemented on the reduced city-level system rather than on a full set of

sector-by-sector moment conditions.

Implication for estimation. An immediate implication of Lemma 1 is that the empirical wage equations do not require separate observations on factory-gate prices p_{djt} . Those prices matter for the model and for the equilibrium allocation across industries, but in the city-level representation they enter only through the composite residuals Ξ_{Hjt} and Ξ_{Ljt} . This is why the GMM can be implemented at the city×time level even though the underlying quantitative model remains multi-industry.

Remark. Lemma 1 concerns the city-level static wage equations and does not justify applying a city-composite diffusion law to Ξ_{kjt} . Persistence estimation instead uses the positive city–sector productivity panel $\exp(\xi_{kdjt})$ and estimates, separately by skill,

$$\xi_{kdjt} - \log\left(\frac{1}{N} \sum_s \exp(\xi_{kds,t-1})\right) = \varrho_k \left[\xi_{kdj,t-1} - \log\left(\frac{1}{N} \sum_s \exp(\xi_{kds,t-1})\right) \right] + \varepsilon_{kdjt}.$$

Sector–transition fixed effects absorb common scale changes. The equality test is performed before the two persistence estimates are used in the quantitative model, and the realized residual ε_{kdjt} is then fixed across counterfactuals.

C.3 Additional Details on Measuring Skill-Biased Technical Change and Core Parameters

This subsection provides additional details on (i) the construction of the skill-biased technical change (SBTC) shifters S_{kjt} used in the calibration and counterfactuals, and (ii) the empirical identification and parameter estimates that discipline the quantitative model. Figure C.1 plots the time evolution of the SBTC exposure measures ΔS_{Hjt} and ΔS_{Ljt} as cross-MSA averages within each year/decade. The figure is intended as a descriptive summary of how routine-task exposure evolves over the sample for highly skilled and less skilled workers. In particular, the series can be positive in the earlier decades and then decline after 1980, which helps motivate the timing of the post-1980 change in the model’s SBTC wedge. Table C.2 reports the first-stage regressions linking the Bartik-style instrumental variable to the endogenous SBTC exposure measures (separately for highly skilled and less skilled workers). The table documents instrument relevance across decades. Identification additionally assumes that the instruments affect the wage equations only through SBTC exposure. The first-stage estimates assess instrument relevance. Table C.3 combines the GMM estimates of the static structural parameters with persistence estimates based on the model’s productivity transition. Panels A

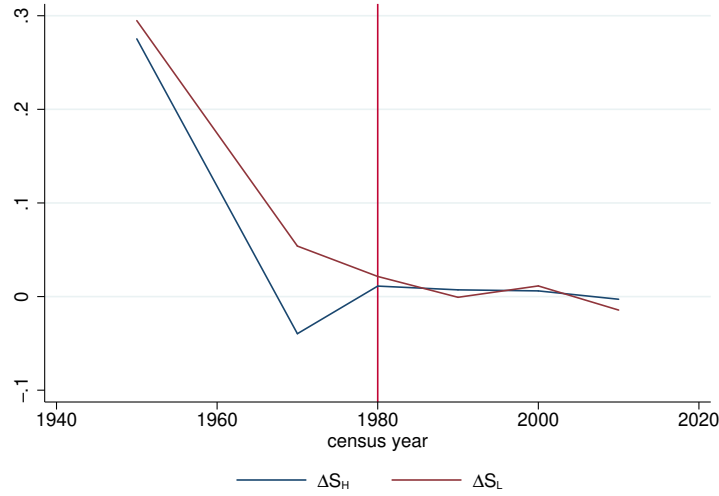
Table C.1: Description of Industries Used for Estimation

Sector	Description	Model Trade Partition
060	Construction	Non-tradable
100-222	Nondurable Manufacturing	Tradable
230-301	Durable Manufacturing	Tradable
310-432	Machinery and Transportation	Non-tradable
440-442	Communication	Non-tradable
450-472	Utilities	Non-tradable
500-532	Wholesale Trade	Tradable
540-571	Nondurable Wholesale Trade	Tradable
580-691	Retail Trade	Non-tradable
700-712	Finance, Insurance, and Real Estate	Tradable
721-752	Business and Repair Services	Tradable
760-791	Personal Services	Non-tradable
800-810	Entertainment and Recreational Services	Non-tradable
812-893	Professional and Related Services	Tradable

Note: Table C.1 lists the industries used for estimation and calibration, and for constructing the ΔS_H and ΔS_L measures, using the 1990 Census Bureau industrial classification applied consistently over time. The model-trade-partition column follows the model's trade partition used in the trade-with-ROW block: tradable sectors are Nondurable Manufacturing, Durable Manufacturing, Wholesale Trade, Nondurable Wholesale Trade, Finance, Insurance, and Real Estate, Business and Repair Services, and Professional and Related Services; all remaining sectors are classified as non-tradable. The classification applies to the model's broad composite sectors; individual Census industries within each sector may differ in tradability.

and B report the five nested GMM specifications. Panel C estimates ϱ_H and ϱ_L separately on transitions ending in 1970, 1980, and 1990 using the decadal unweighted global-mean transition; the equality test rejects a common persistence parameter, so the quantitative model uses skill-specific estimates. The quantitative baseline restricts the statistically insignificant γ^L and ϕ^L to zero. The table reports structural parameters; coefficients on the SBTC proxies enter only the first-step wage regressions.

Figure C.1: Aggregate evolution of ΔS_{Hjt} and ΔS_{Ljt}



Note: Figure C.1 reports the 10-year changes in the two series. Each point is the unweighted cross-MSA average of the city-level measures ΔS_{Hjt} and ΔS_{Ljt} within that period. The legend labels these aggregates as ΔS_H and ΔS_L for compactness.

Table C.2: First-Stage Estimates of Models for Routine Occupation Share Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
$\Delta \hat{Z}_{j,t-10}^H$	3.574*** (0.746)	4.054*** (1.112)	3.835*** (0.715)	5.562*** (1.214)	3.579*** (0.778)	3.237*** (0.777)
F-stat (excl. instr.)	22.95	13.29	28.77	20.97	21.15	17.36
N	125	119	270	249	283	283
Adj. R^2	0.150	0.0943	0.0936	0.0745	0.0667	0.0548
Panel B						
$\Delta \hat{Z}_{j,t-10}^L$	1.146*** (0.393)	0.946** (0.372)	0.948*** (0.315)	2.729*** (0.582)	2.855*** (0.579)	2.986*** (0.661)
F-stat (excl. instr.)	8.510	6.483	9.056	21.98	24.34	20.41
N	144	119	270	249	283	283
Adj. R^2	0.0499	0.0444	0.0291	0.0780	0.0765	0.0644

Note: Table C.2 reports estimates from regressions of the endogenous SBTC exposure measures ΔS_{kjt} on the Bartik-style instruments $\Delta \hat{Z}_{kjt}$. Standard errors are in brackets. Column (1) reports the 1950 cross section and columns (2)–(6) report 1970–2010. The sample does not contain 1960. Panel A reports the results for highly skilled workers and Panel B for less skilled workers. The table also reports the excluded-instrument F-statistic. Asterisks ***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table C.3: Model Estimates for 1940–2010

	A	B	C	D	E
Panel A: Labor Demand					
ρ	0.60*** (0.23)	0.75*** (0.18)	0.53*** (0.17)	0.69*** (0.19)	0.67*** (0.19)
γ^H	0.72*** (0.28)	0.55* (0.28)	0.72*** (0.25)		0.53** (0.26)
γ^L	0.07 (0.24)	0.06 (0.26)	0.12 (0.22)		0.11 (0.22)
ϕ^H		0.27*** (0.08)			0.25*** (0.08)
ϕ^L		0.21 (0.13)			0.21 (0.14)
Panel B: Labor Supply					
θ	0.62*** (0.10)	0.61*** (0.10)	0.43*** (0.12)	0.42*** (0.10)	0.42*** (0.12)
γ^p			0.91*** (0.12)		0.93*** (0.12)
Panel C: Productivity Persistence					
ϱ_H					0.378*** (0.036)
ϱ_L					0.703*** (0.034)
$\chi^2(1) : \varrho_H = \varrho_L$					63.47
p -value					< 0.001
Persistence observations per skill					2,940
GMM observations	705	705	705	705	705

Note: Panels A and B report the static GMM estimates for 1940–2010. Columns (A)–(E) are: (A) skill spillovers only; (B) skill and size spillovers; (C) skill spillovers with endogenous amenities but no size spillovers; (D) no production or amenity spillovers; and (E) the preferred unrestricted static specification. The quantitative baseline carries column (E)’s H coefficients but imposes $\gamma^L = \phi^L = 0$ because both L estimates are statistically insignificant. Panel C is estimated separately on the positive productivity panel using transitions ending in 1970, 1980, and 1990, with sector–transition fixed effects, 70 city clusters, and 2,940 observations per skill. Its operator is the decadal unweighted arithmetic global mean in productivity levels. The displayed standard errors are clustered by city. The equality test uses the estimated cross-skill covariance. No 2000 or 2010 holdout observation enters the persistence estimates. Asterisks ***, **, and * denote significance at the 0.01, 0.05, and 0.10 levels.

C.4 Validation of Amenities and Destination Components of Bilateral Migration Costs

This section assesses the economic interpretation of the recovered amenities and the destination components of bilateral migration costs by comparing them with independent proxies of local quality of life and migration frictions. The recovered time-varying amenities, A_{kjt} , are interpreted as the residual utility value of living in location j for workers of skill group k ,

Table C.4: Aggregate Wage Growth: Data and Model

Outcome	Observed panel	Model
Highly skilled wage growth, 1980–2010	0.217	0.217
Less skilled wage growth, 1980–2010	-0.109	-0.059
Log skill-premium change, 1980–2010	0.326	0.276
Log skill premium, 2010	0.707	0.658

Note: Both columns cover the same 70 MSAs in 1980 and 2010. The observed column uses city–skill wage indices and employment from the estimation panel; the model column uses the quantitative baseline. In each column, aggregate wages are skill-specific employment-weighted means of city wage levels. Wage growth is the change in their logarithm, and the log skill premium is the difference between the two aggregate log wages. The 1980 reporting normalization sets the model premium to its observed value in this sample, leaving growth unchanged. This is a within-sample comparison using the wage panel that enters the baseline reconstruction.

after accounting for wages, prices, and housing costs. The fixed time-invariant terms $m_{k2}(j)$ are the destination components of bilateral migration costs and summarize persistent entry costs that reduce inflows into location j .

Destination Components of Bilateral Migration Costs. Following [Desmet, Nagy and Rossi-Hansberg \(2018\)](#), I refer to $m_{k2}(j)$ as the destination component of the bilateral migration cost. This component may encompass housing constraints, regulatory barriers, and other impediments to mobility, consistent with the broader interpretations in [Hsieh and Moretti \(2019\)](#) and [Eeckhout, Pinheiro and Schmidheiny \(2014\)](#). I correlate $\log m_{k2}(j)$ with log rents, the housing-supply-elasticity measure of [Saiz \(2010\)](#), the Wharton Land Regulation Index ([Gyourko, Saiz and Summers, 2008](#)), and changes in college and non-college population shares. Table C.5 reports unweighted Pearson correlations using the same available city observations for both skill groups, with the number of observations shown for each comparison.

The largest correlations are with log rents: 0.444 for highly skilled workers and 0.323 for less skilled workers. The Wharton correlations are smaller but positive (0.161 and 0.089), while the Saiz housing-supply correlations are near zero and slightly negative (-0.013 and -0.028). Correlations with both changes in skill shares are small and positive. The largest destination components occur in New York, Los Angeles, Chicago, and Philadelphia for both skill groups. These correlations provide descriptive evidence about the recovered destination costs; rents also enter the inversion used to recover them.

Amenities. Following [Diamond \(2016\)](#) and [Gyourko, Mayer and Sinai \(2013\)](#), I compare the recovered 2010 amenity residuals A_{kjt} with long-run temperature, reversed EPA Air Quality

Index, reversed FBI violent-crime rates, and cultural establishments per capita. Table C.6 reports unweighted Pearson correlations and the number of city observations available for each comparison.

The proxy data provide six matching observations in the 70-city model sample. Missing values limit individual correlations to three to six observations. All four displayed relationships are positive for both skill groups: temperature is the smallest, while cleaner air, lower violent crime, and cultural establishments have larger positive sample correlations. The reversed-crime correlations are 0.933 for highly skilled workers and 0.950 for less skilled workers, with three observations for each correlation. These small samples limit the precision of the estimates and their applicability to the broader model sample.

Discussion. The proxy correlations and rankings among the 70 model metropolitan areas help describe the recovered amenities and destination components of migration costs. The destination components of bilateral migration costs are most positively associated with rents and more weakly with land-use regulation; the housing-supply correlation is essentially zero. The amenity correlations are positive in the available observations. These comparisons supplement the descriptive interpretation of the recovered amenities and migration costs; they enter neither the calibration nor the recovery of these quantities.

Table C.5: Validation of Destination Components of Bilateral Migration Costs

Variable	Correlation with $\log m_{k2}(j)$		
	Highly Skilled	Less Skilled	Pair count
log Average Rent Level	0.444	0.323	70
Housing Supply Elasticity (Saiz, 2010)	-0.013	-0.028	68
Land Regulation Index (Wharton)	0.161	0.089	69
Δ college share (Δ col)	0.136	0.132	70
Δ no-college share (Δ no-col)	0.114	0.091	70

Note: Table C.5 reports unweighted Pearson correlations of the logged, time-invariant destination components of bilateral migration costs, $m_{H2}(j)$ and $m_{L2}(j)$, with 2010 metropolitan proxies. All 70 model metropolitan areas join uniquely; pair counts vary with proxy availability. H denotes at least a college degree and L denotes less than college.

Table C.6: Validation of Inverted Amenities

Variable	Correlation with 2010 amenities		Pair count
	A_H	A_L	
Average Temperature (NOAA)	0.377	0.209	5
Air Quality Index (EPA, reversed)	0.855	0.861	6
Violent Crime Rate (FBI, reversed)	0.933	0.950	3
Cultural Establishments per Capita	0.763	0.827	6

Note: Table C.6 reports unweighted Pearson correlations of the sealed 2010 amenities with metropolitan quality-of-life proxies. Of the 18 proxy rows, only six overlap the 70-area model universe, and missing proxy values reduce the row-specific pair counts further. Reversed AQI and crime are coded so higher values mean cleaner air and lower crime. These small-sample correlations provide descriptive validation evidence. H denotes at least a college degree and L denotes less than college.

Table C.7: Top and Bottom 10 Metropolitan Areas by Inferred Amenities, 2010

Highly Skilled		Less Skilled	
Metropolitan Area	Rank	Metropolitan Area	Rank
<i>Panel A. Top 10 Amenities</i>			
New York, NY–Northeastern NJ	1	New York, NY–Northeastern NJ	1
Los Angeles–Long Beach, CA	2	Los Angeles–Long Beach, CA	2
Chicago, IL	3	Chicago, IL	3
Dallas–Fort Worth, TX	4	Dallas–Fort Worth, TX	4
Washington, DC–MD–VA	5	Houston–Brazoria, TX	5
Houston–Brazoria, TX	6	Riverside–San Bernardino, CA	6
Philadelphia, PA–NJ	7	Philadelphia, PA–NJ	7
San Francisco–Oakland–Vallejo, CA	8	Atlanta, GA	8
Atlanta, GA	9	Detroit, MI	9
Detroit, MI	10	Phoenix, AZ	10
<i>Panel B. Bottom 10 Amenities</i>			
Flint, MI	1	Worcester, MA	1
South Bend–Mishawaka, IN	2	South Bend–Mishawaka, IN	2
Worcester, MA	3	Trenton, NJ	3
Trenton, NJ	4	Flint, MI	4
Beaumont–Port Arthur–Orange, TX	5	Madison, WI	5
Santa Barbara–Santa Maria–Lompoc, CA	6	Santa Barbara–Santa Maria–Lompoc, CA	6
Spokane, WA	7	Des Moines, IA	7
Des Moines, IA	8	Spokane, WA	8
Madison, WI	9	Springfield–Holyoke–Chicopee, MA	9
Youngstown–Warren, OH–PA	10	Columbia, SC	10

Note: Table C.7 ranks the exact sealed 2010 amenity inputs for highly skilled and less skilled workers. The inversion is mean-zero within skill and date and no smoothing is applied. Ranks are relative across the 70 metropolitan areas used by the model; metro code breaks ties.

Table C.8: Top and Bottom 10 Metropolitan Areas by Destination Components of Bilateral Migration Costs, 2010

Highly Skilled		Less Skilled	
Metropolitan Area	Rank	Metropolitan Area	Rank
<i>Panel A. Largest 10 Destination Components of Bilateral Migration Costs</i>			
New York, NY–Northeastern NJ	1	New York, NY–Northeastern NJ	1
Los Angeles–Long Beach, CA	2	Los Angeles–Long Beach, CA	2
Chicago, IL	3	Chicago, IL	3
Philadelphia, PA–NJ	4	Philadelphia, PA–NJ	4
Washington, DC–MD–VA	5	Detroit, MI	5
San Francisco–Oakland–Vallejo, CA	6	Dallas–Fort Worth, TX	6
Detroit, MI	7	Houston–Brazoria, TX	7
Dallas–Fort Worth, TX	8	San Francisco–Oakland–Vallejo, CA	8
Boston, MA–NH	9	St. Louis, MO–IL	9
Houston–Brazoria, TX	10	Baltimore, MD	10
<i>Panel B. Smallest 10 Destination Components of Bilateral Migration Costs</i>			
Santa Barbara–Santa Maria–Lompoc, CA	1	Madison, WI	1
Trenton, NJ	2	Santa Barbara–Santa Maria–Lompoc, CA	2
South Bend–Mishawaka, IN	3	Trenton, NJ	3
Madison, WI	4	Worcester, MA	4
Worcester, MA	5	South Bend–Mishawaka, IN	5
Spokane, WA	6	Des Moines, IA	6
Flint, MI	7	Spokane, WA	7
Beaumont–Port Arthur–Orange, TX	8	Albuquerque, NM	8
Des Moines, IA	9	Columbia, SC	9
Albuquerque, NM	10	Tucson, AZ	10

Note: Table C.8 ranks the exact time-invariant destination components of bilateral migration costs for highly skilled and less skilled workers. The displayed 2010 ranking is relative across the 70 metropolitan areas used by the model; metro code breaks ties.

C.5 Alternative Approaches to Migration Costs

A well-known study by [Kennan and Walker \(2011\)](#) provides estimates of large moving costs from one state to another in the U.S. but does not distinguish between highly skilled and less skilled workers. By contrast, [Notowidigdo \(2020\)](#) provides separate migration costs for highly skilled and less skilled workers, closely related to the migration costs in my setting. This makes his estimates applicable to my setting. Specifically, [Notowidigdo \(2020\)](#) uses an exponential function to estimate migration costs. The functional form he estimates is as follows:

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

where x_j denotes MSA characteristics such as population. Despite having only two parameters, this functional form allows migration costs to vary nonlinearly with those characteristics, as discussed in [Notowidigdo \(2020\)](#).

The estimated parameters of the exponential migration-cost function differ between highly skilled and less skilled workers. [Notowidigdo \(2020\)](#) identifies them using Bartik instruments. A parsimonious benchmark motivated by those estimates would normalize the highly skilled destination component of the bilateral migration cost to one and set the less-skilled component uniformly to 1.16 across MSAs, rather than feed the full city-varying exponential schedule into the model. This alternative parameterization is not used in the baseline or the reported counterfactuals; it is included only to describe a possible methodological benchmark. Applying it quantitatively would also require accounting for differences between the two model structures.

C.6 Productivity, Wage, and Amenity Inversion

This subsection gives the complete sequence used to recover productivity, the cross-city wage residuals, and amenities. The inversion is performed at the six observed dates $\mathcal{T} = \{1960, 1970, 1980, 1990, 2000, 2010\}$ on the common positive-support city–skill–sector panel. The inputs are city wages (W_{Hjt}, W_{Ljt}) , city employment (H_{jt}, L_{jt}) , sectoral employment (H_{djt}, L_{djt}) , bilateral trade costs, and the parameters estimated or calibrated before the inversion.

Effective productivity. Let $K_{Hdj} = H_{dj}$ and $K_{Ldj} = L_{dj}$. At each sector–date pair, origin revenue and destination absorption are

$$\mathcal{R}_{dj} = W_{Hjt}H_{dj} + W_{Ljt}L_{dj}, \quad E_{jt} = W_{Hjt}H_{jt} + W_{Ljt}L_{jt}.$$

Define their national shares by $r_{djt} = \mathcal{R}_{djt} / \sum_s \mathcal{R}_{dst}$ and $e_{jt} = E_{jt} / \sum_s E_{st}$. Relative producer prices solve the Armington system

$$\pi_{jj't}^d = \frac{(\tau_{jj'}^d p_{djt})^{1-\sigma_d}}{\sum_s (\tau_{sj'}^d p_{dst})^{1-\sigma_d}}, \quad r_{djt} = \sum_{j'} \pi_{jj't}^d e_{j't},$$

with the sector-date normalization $\sum_j \log p_{djt} = 0$. This normalization removes the arbitrary common price scale but preserves all cross-city relative prices. Perfect competition implies $Y_{djt} = \mathcal{R}_{djt} / p_{djt}$. The sectoral CES first-order condition can then be written as

$$W_{kjt} = p_{djt} \eta_{kdjt}^{obs} K_{kdjt}^{\rho-1} Y_{djt}^{1-\rho},$$

so effective productivity is recovered cell by cell as

$$\eta_{kdjt}^{obs} = \left(\frac{K_{kdjt}}{Y_{djt}} \right)^{1-\rho} \frac{W_{kjt}}{p_{djt}}.$$

Strictly positive employment, wages, prices, and output on the common support ensure that the inversion is finite and positive. Substituting the recovered η_{kdjt}^{obs} into the first-order conditions reproduces the sectoral factor allocations used in this step.

Productivity, persistence, and realized innovations. Write the agglomeration component in equation (18) as

$$G_{Hjt} = \left(\frac{H_{jt}}{L_{jt}} \right)^{\gamma^H} (H_{jt} + L_{jt})^{\phi^H}, \quad G_{Ljt} = 1.$$

After removing this term and the original SBTC mapping, the positive productivity level supplied to the quantitative model is

$$\exp(\xi_{kdjt}) = \frac{\eta_{kdjt}^{obs}}{G_{kjt} S_{kjt}^{\Lambda_{kt}}}.$$

Persistence is estimated once from the sector-normalized productivity profiles obtained by removing the estimated static GMM components from η_{kdjt}^{obs} . For each skill, let $\tilde{\xi}_{kdjt}$ denote the log of that profile. Its unweighted arithmetic city mean in levels is $N^{-1} \sum_s \exp(\tilde{\xi}_{kds,t-1})$. The estimating equation is

$$\tilde{\xi}_{kdjt} - \log \left(\frac{1}{N} \sum_s \exp(\tilde{\xi}_{kds,t-1}) \right) = a_{kdt} + \varrho_k \left[\tilde{\xi}_{kdj,t-1} - \log \left(\frac{1}{N} \sum_s \exp(\tilde{\xi}_{kds,t-1}) \right) \right] + u_{kdjt}.$$

The sample contains the decadal transitions ending in 1970, 1980, and 1990. The fixed effects a_{kdt} absorb sector–transition means, and inference is clustered by city. I estimate ϱ_H and ϱ_L separately and test their equality rather than impose it.

Once these persistence estimates are fixed, the six-date log-productivity path used by the model is connected by

$$\xi_{kdjt} = \varrho_k \xi_{kdj,t-1} + (1 - \varrho_k) \log \left(\frac{1}{N} \sum_s \exp(\xi_{kds,t-1}) \right) + \varepsilon_{kdjt}.$$

The realized innovation is therefore the identity residual

$$\varepsilon_{kdjt} = \xi_{kdjt} - \varrho_k \xi_{kdj,t-1} - (1 - \varrho_k) \log \left(\frac{1}{N} \sum_s \exp(\xi_{kds,t-1}) \right).$$

It records the part of the observed change in productivity geography that is not generated by the estimated diffusion component.

Centered-wage residual profiles. The effective-productivity inversion above uses sectoral factor demand. Recombining those productivities with the complete trade, housing, and labor-market system does not by itself reproduce the observed cross-city wage profiles. I therefore recover one city profile b_{kjt} for each skill and observed date, common across sectors. It enters effective productivity multiplicatively as $\exp(b_{kjt})$.

Let $\omega_{kjt} = E_{kjt}^{obs}/E_{kt}^{tot}$ be the observed city employment share of skill k . Observed and model-centered log wages are

$$\tilde{w}_{kjt}^{obs} = \log W_{kjt}^{obs} - \sum_s \omega_{kst} \log W_{kst}^{obs}, \quad \tilde{w}_{kjt}^{(r)} = \log W_{kjt}^{(r)} - \sum_s \omega_{kst} \log W_{kst}^{(r)}.$$

Starting from $b_{kjt}^{(0)} = 0$, I hold the city allocations at their observed values, solve the remaining equilibrium conditions, and update

$$b_{kjt}^{(r+1)} = b_{kjt}^{(r)} + \frac{1}{2} \left(\tilde{w}_{kjt}^{obs} - \tilde{w}_{kjt}^{(r)} \right).$$

After every update I subtract the unweighted city mean, so $\sum_j b_{kjt}^{(r+1)} = 0$. The iteration stops when the largest absolute centered-log-wage discrepancy is below 10^{-3} and the equilibrium residual is below 10^{-8} . With 70 cities, two skills, and six dates, the normalization leaves $2 \times 6 \times (70 - 1) = 828$ residual coefficients. They are not structural parameters, and no national skill–date intercept is added.

Amenity inversion. The inversion uses the normalized preference index and destination components in equations (15)–(16). Amenities are recovered only after wages, rents, and prices have been determined at the observed allocations. Let $s_{kjt} = E_{kjt}/E_{kt}^{tot}$ and define the part of the location-choice index excluding amenities and the fixed destination component of the bilateral migration cost by

$$\tilde{\delta}_{kjt} = \log(W_{kjt} - R_{jt}\bar{O}) - \theta \log P_{jt} - (1 - \theta) \log R_{jt} + \gamma^p \log(H_{jt}/L_{jt}).$$

The model share implied by a candidate amenity vector is

$$\hat{s}_{kjt} = \frac{\exp\{\tilde{\delta}_{kjt} + A_{kjt} - \log m_{k2}(j)\}}{\sum_r \exp\{\tilde{\delta}_{krt} + A_{krt} - \log m_{k2}(r)\}}.$$

Hence the closed-form inverse is

$$q_{kjt} = \log s_{kjt} - \tilde{\delta}_{kjt} + \log m_{k2}(j), \quad A_{kjt} = q_{kjt} - \frac{1}{N} \sum_{r=1}^N q_{krt}.$$

This gives $\sum_j A_{kjt} = 0$ and $\hat{s}_{kjt} = s_{kjt}$. The inversion is performed exactly once for each skill and date, producing 12 amenity vectors. The numerical checks require the largest share discrepancy and the normalization residual to be below 10^{-10} .

What the inversion determines. Observed city allocations enter directly, amenities reproduce their shares, and the residual profiles reproduce centered wages within the stated tolerance. Cross-city wage dispersion and convergence coefficients follow from the matched wage profiles. Because the b_{kjt} profiles have mean zero and contain no national intercepts, national skill–date wage levels, aggregate wage growth, and skill-premium levels remain determined by the aggregate calibration and the equilibrium system. The log productivities and realized innovations, the centered-wage residual profiles, and the amenity vectors are held fixed in every counterfactual; only the structural component named in the exercise is changed. No smoothing is applied between or within the six observed dates.

Table C.9 summarizes only the 1960 cross-city component of the productivity input using the arithmetic mean across the same 14 sectors supplied to the model. The table describes the model’s initial productivity distribution.

Table C.9: Top and Bottom 10 Metropolitan Areas by Inferred Productivity, 1960

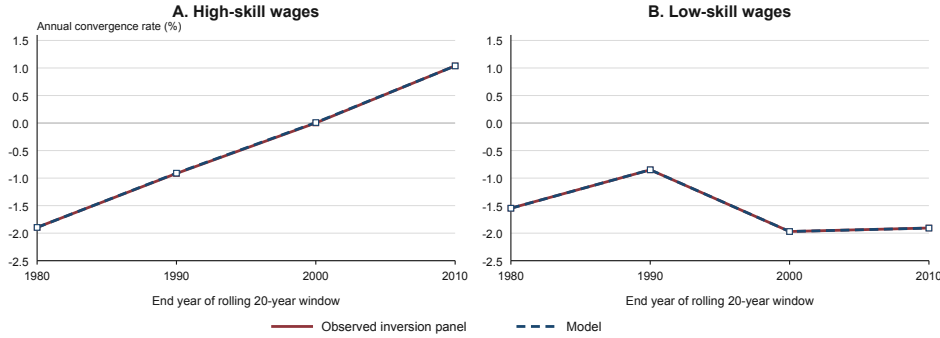
Highly Skilled		Less Skilled	
Metropolitan Area	Rank	Metropolitan Area	Rank
<i>Panel A. Top 10 Productivity</i>			
Beaumont–Port Arthur–Orange, TX	1	Flint, MI	1
Grand Rapids, MI	2	South Bend–Mishawaka, IN	2
Charlotte–Gastonia–Rock Hill, NC–SC	3	Grand Rapids, MI	3
Flint, MI	4	Toledo, OH–MI	4
South Bend–Mishawaka, IN	5	Des Moines, IA	5
Birmingham, AL	6	Beaumont–Port Arthur–Orange, TX	6
Trenton, NJ	7	Youngstown–Warren, OH–PA	7
Worcester, MA	8	Worcester, MA	8
Toledo, OH–MI	9	Santa Barbara–Santa Maria–Lompoc, CA	9
Indianapolis, IN	10	Springfield–Holyoke–Chicopee, MA	10
<i>Panel B. Bottom 10 Productivity</i>			
New York, NY–Northeastern NJ	1	New York, NY–Northeastern NJ	1
Washington, DC–MD–VA	2	Philadelphia, PA–NJ	2
Los Angeles–Long Beach, CA	3	Washington, DC–MD–VA	3
San Francisco–Oakland–Vallejo, CA	4	Los Angeles–Long Beach, CA	4
Chicago, IL	5	Boston, MA–NH	5
Philadelphia, PA–NJ	6	Chicago, IL	6
Boston, MA–NH	7	Columbia, SC	7
Denver–Boulder, CO	8	Dallas–Fort Worth, TX	8
Miami–Hialeah, FL	9	Tampa–St. Petersburg–Clearwater, FL	9
Riverside–San Bernardino, CA	10	Miami–Hialeah, FL	10

Note: Table C.9 ranks the 1960 baseline productivity levels $\exp(\xi_{kdj,1960})$ supplied to the model for highly skilled and less skilled workers. City productivity is the simple mean of the 14 sector-specific components.

The full quantitative input also contains realized innovations and the mean-zero productivity residual described in the text.

C.7 Spatial Moments in the model and Data

Figure C.2: Model and Data: Regional Convergence



Note: Figure C.2 plots raw weighted 20-year convergence coefficients for the four common observed windows. The x-axis gives the window end year. Each coefficient uses the corresponding skill group's employment in the endpoint year as weights. The connecting lines are visual guides between the four reported coefficients and do not represent additional model states.

Table C.10: Model and Data: Additional Spatial Outcomes

	Highly skilled secular migration	Less skilled secular migration	Change in β^{HL}	Change in δ
Model	-0.61 pp	-0.37 pp	0.37 pp	0.20
Data	-0.60 pp	-0.37 pp	0.37 pp	0.19

Note: Model and data use the same definitions, cities, and years. The first two columns compare averages for transition end-years 1961–1980 and 1990–2010. Observed city employment is interpolated log-linearly between decadal inputs for this calculation. The third column compares 1960–1980 and 1990–2010 slopes of annualized 20-year log skill-ratio growth on its initial log level, weighted by initial total employment; observed employment is normalized to the model's national skill totals. The fourth compares the unweighted log-premium/log-skill-ratio correlations in 1980 and 2010. Secular migration and convergence changes are in percentage points; correlation changes are in levels. Model and data are reported separately.

D Complementary Explanations

This section evaluates complementary mechanisms and quantitative sensitivities for the reversal in regional wage convergence. It starts with physical capital channels—including transition dynamics and capital–skill complementarity—in Section D.1, followed by human capital accumulation in Section D.2. It then analyzes trade openness in Section D.3. The remaining subsections cover alternative definitions of local SBTC (Section D.4), housing regulation (Section D.5), structural transformation and firm sorting (Section D.6), and *Right-to-Work* laws and unions (Section D.7). Finally, Section D.8 reports the whole-horizon no-diffusion benchmark. Among the individual mechanism counterfactuals in Table 1, the full No Agglomeration shutdown produces the largest reduction in the high-skill convergence reversal, but does not eliminate it; No SBTC alone changes the reversal only modestly; and their joint shutdown is slightly more than additive for highly skilled wages. The alternative SBTC paths yield outcomes close to the baseline. By contrast, No Diffusion produces extreme skill-specific paths, and the capital extension generates stronger high-skill convergence and weaker low-skill convergence. These two cases show that the quantitative results are sensitive to diffusion and physical capital. All comparisons hold the realized productivity innovations, productivity residuals, and amenity path fixed.

D.1 Physical Capital, Complementarity, and Transition Dynamics

Physical-capital transition dynamics could affect regional convergence, as emphasized by Kleinman, Liu and Redding (2023) for U.S. states. Faster capital deepening in initially poorer locations would tend to promote convergence and could therefore offset, rather than generate, the post-1980 slowdown. To examine whether the skill-specific pattern also appears at a broader geographic scale, I repeat the regressions in Table A.2 using states rather than MSAs. Evidence at the state level can reveal whether broad regional forces accompany the metropolitan pattern, although it cannot by itself rule out localized agglomeration within states. Table D.1 shows a decline in aggregate convergence in Panel A, as in the MSA-level analysis, but of smaller magnitude. Panel B shows that highly skilled wages across states continue to converge at -1.1% per year, in contrast to the MSA-level results. This pattern is consistent with the state-level evidence in Kleinman, Liu and Redding (2023) and motivates the paper’s focus on the localized interaction between SBTC and agglomeration as its main explanation for the skill-specific metropolitan divergence.

D.1.1 Model with Physical Capital and Transition Dynamics

The baseline quantitative model pins down the endogenous path of spatial outcomes by solving a sequence of within-period spatial equilibria given the initial technology distribution and the exogenous evolution of fundamentals (including skill-biased technical change and trade/migration costs). A natural concern is that this approach abstracts from additional transition dynamics implied by predetermined physical capital: because physical capital adjusts gradually, locations may temporarily deviate from their long-run allocations.

To address this concern, this appendix introduces a tractable extension with *location-specific physical capital* and forward-looking investment. The extension follows the dynamic spatial general equilibrium framework of [Kleinman, Liu and Redding \(2023\)](#), in which (i) capital is immobile once installed, (ii) physical capital is owned by immobile local landlords, and (iii) landlords choose consumption and investment subject to a standard capital accumulation equation. The purpose is to quantify whether the additional transition dynamics meaningfully alter the key convergence and migration responses in my counterfactual exercises.

Environment. Time is discrete, $t = 0, 1, \dots, T$. The economy consists of a set of locations $i \in \{1, \dots, N\}$ and industries $d \in \{1, \dots, D\}$. Workers are of two types $s \in \{H, L\}$ (highly skilled and less skilled). Let $n_{s,it}$ denote the mass of type- s workers living in location i at t , and let $w_{s,it}$ denote the corresponding wage. Each location has an endogenous housing price (rent) R_{it} and a price index for the tradable consumption composite P_{it} . All preferences, trade structure, migration costs, amenities, and the evolution of productivity/skill-bias shifters are as in the baseline model, except where explicitly modified below.

Production with physical capital. Each location-industry pair (i, d) produces a differentiated variety under perfect competition using highly skilled and less skilled labor together with installed capital. Value added in (i, d, t) is

$$Y_{idt} = Z_{idt} K_{idt}^{1-\mu_K} \left(\mathcal{L}_{idt} \right)^{\mu_K}, \quad (38)$$

where $\mu_K \in (0, 1)$ is the labor share and Z_{idt} is an exogenous productivity term (capturing the same fundamental shocks as in the baseline model). The labor aggregate is a CES combination of skill inputs,

$$\mathcal{L}_{idt} = \left(A_{L,idt} L_{idt}^\rho + A_{H,idt} H_{idt}^\rho \right)^{1/\rho}, \quad (39)$$

with $\rho < 1$ and where $(A_{H,idt}, A_{L,idt})$ embed the baseline agglomeration terms and the evolution of skill-biased productivity shifters.

Physical capital is *location-specific*. For tractability in the multi-industry environment, I allocate location-level capital across industries using fixed nonnegative shares:

$$K_{idt} = \omega_d K_{it}, \quad \sum_{d=1}^D \omega_d = 1, \quad \omega_d \geq 0. \quad (40)$$

When $\omega_d = 1/D$, this corresponds to an equal split across industries. In the implementation, $\omega_d = 1/D$ and the expenditure shares also satisfy $\alpha_d = 1/D$; the two vectors are numerically equal but represent distinct objects.

Factor payments. Let r_{it}^K denote the rental rate on physical capital in location i . Competitive factor markets imply the standard Cobb–Douglas income shares at the location level:

$$\sum_d (w_{H,it} H_{idt} + w_{L,it} L_{idt}) = \mu_K \sum_d p_{idt} Y_{idt}, \quad r_{it}^K K_{it} = (1 - \mu_K) \sum_d p_{idt} Y_{idt}, \quad (41)$$

so that

$$r_{it}^K K_{it} = \frac{1 - \mu_K}{\mu_K} (w_{H,it} n_{H,it} + w_{L,it} n_{L,it}), \quad (42)$$

where I use the fact that labor is mobile across industries within a location and the wage bill equals $w_{H,it} n_{H,it} + w_{L,it} n_{L,it}$.

Local landlords and capital accumulation. Following [Kleinman, Liu and Redding \(2023\)](#), each location has a representative immobile landlord who owns physical capital K_{it} . The landlord chooses consumption c_{it}^K and next-period physical capital $K_{i,t+1}$. Investment uses the same tradable composite as consumption (price index P_{it}). The landlord's period budget constraint is

$$r_{it}^K K_{it} = P_{it} (c_{it}^K + K_{i,t+1} - (1 - \delta) K_{it}), \quad (43)$$

where $\delta \in [0, 1)$ is the depreciation rate. Define the gross return on physical capital (in units of the tradable composite)

$$\mathcal{R}_{it} \equiv (1 - \delta) + \frac{r_{it}^K}{P_{it}}. \quad (44)$$

The landlord has CRRA preferences $\sum_{t \geq 0} \beta_K^t u(c_{it}^K)$ with intertemporal elasticity of substitution ψ . As shown in [Kleinman, Liu and Redding \(2023\)](#), the optimal policy takes the form of a saving rule

$$c_{it}^K = \varsigma_{it} \mathcal{R}_{it} K_{it}, \quad K_{i,t+1} = (1 - \varsigma_{it}) \mathcal{R}_{it} K_{it}, \quad (45)$$

where ς_{it} depends on expected future returns.

Log-utility implementation. To keep the robustness exercise computationally tractable at a fine geographic resolution, I adopt the log-utility special case $\psi = 1$. In this case, the saving rate is constant and equal to the discount factor:

$$K_{i,t+1} = \beta_K \mathcal{R}_{it} K_{it}. \quad (46)$$

Combining (42)–(46) yields a simple transition equation for K_{it} given the within-period equilibrium objects $(w_{H,it}, w_{L,it}, n_{H,it}, n_{L,it}, P_{it})$.

Goods markets with capital income. All workers are hand-to-mouth as in the baseline model, so worker tradable expenditure is labor income net of Stone–Geary housing expenditure. Local owners allocate capital income between consumption and net investment according to (46). Total nominal factor income in location i is

$$\mathcal{Y}_{it} = \left(w_{H,it} n_{H,it} + w_{L,it} n_{L,it} \right) + r_{it}^K K_{it} = \frac{1}{\mu_K} \left(w_{H,it} n_{H,it} + w_{L,it} n_{L,it} \right), \quad (47)$$

where the second equality uses (42). The capital extension uses the same goods-market-clearing conditions as the baseline. Domestic absorption equals worker tradable spending plus locally owned housing income and capital income. Producer prices and wages adjust until sectoral producer revenue equals domestic plus ROW demand revenue, both skill-factor markets clear, zero profit and factor-share identities hold, and aggregate ROW exports equal imports.

Equilibrium and computation. A dynamic equilibrium consists of sequences of allocations and prices $\{n_{s,it}, w_{s,it}, R_{it}, P_{it}, K_{it}\}_{t=0}^T$ such that: (i) for a given state $(\{n_{s,it}\}_s, K_{it})$ and the period- t fundamentals, all within-period goods, labor, and housing markets clear and workers’ location choices satisfy the baseline migration block; and (ii) physical capital evolves according to (46).

Algorithm. The computational implementation nests the existing solver:

1. Initialize K_{i0} from data together with the baseline initial conditions.
2. For each period t :
 - (a) Take K_{it} as predetermined.
 - (b) Solve the within-period spatial equilibrium using the same fixed-point routine as in the baseline, except that labor demand and wage updating use the Cobb–Douglas-with-capital production block (38)–(39).

(c) Compute r_{it}^K from (42), update $\mathcal{R}_{it}$ from (44) and update $K_{i,t+1}$ via (46).

Additional parameters and data. Relative to the baseline model, the robustness exercise introduces three scalars $\{\mu_K, \delta, \beta_K\}$ and an initial physical-capital distribution $\{K_{i0}\}_{i=1}^N$.

Initial capital is constructed from the 1960 city-level skill populations, wages, and rent measure in the model’s inversion input. I use the factor-share identity to obtain a provisional city capital distribution and normalize its cross-city mean to one, preserving relative spatial dispersion. This initialization does not use BEA fixed-asset data.

This capital block is adapted from [Kleinman, Liu and Redding \(2023\)](#). The key maintained assumptions are: physical capital is immobile once built, owned by immobile local landlords; investment uses the consumption composite; and the log-utility special case yields a constant saving rate and a closed-form transition for K_{it} . The only additional simplification I impose for tractability in a multi-industry version is the fixed-share allocation in (40). This assumption makes the extension tractable but limits the interpretation of the exercise to the transition dynamics generated under that allocation rule.

Table D.2 shows that including physical capital substantially changes the convergence results. The economy with physical capital generates stronger convergence for highly skilled wages and weaker convergence for less skilled wages, reversing the empirical skill asymmetry. It also leaves a material aggregate low-skill wage-growth gap. The exercise therefore shows that predetermined local capital matters quantitatively, but it does not reproduce the observed post-1980 pattern.

D.1.2 Model with Endogenous Innovation Rate: Capital-Skill Complementarity

Yet physical capital might matter through channels beyond transition dynamics. Baseline SBTC is exogenous and varies by location j . However, I could allow SBTC to be modeled as “technological adoption” following [Beaudry, Doms and Lewis \(2010\)](#). 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 latest technology uses skilled labor more intensively. Specifically, where there is a higher concentration of highly skilled workers, there is also a higher ratio of computers per worker.

The production function with the old technology K is equal to

$$Y_d = K^{(1-\alpha)}[aH^\rho + (1-a)L^\rho]^{\frac{\alpha}{\rho}} \quad (48)$$

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

equal to

$$Y_d = PC^{(1-\alpha)}[bH^\rho + (1-b)L^\rho]^{\frac{\alpha}{\rho}} \quad (49)$$

where $0 < a < b < 1$, and PC denotes the new computer-based technology. Firms choose adoption intensity, which increases with $\frac{H}{L}$. Before the availability of the PC technology, location j that had higher supply of skilled labor also had relatively low wages (because of a congestion effect on skills). Therefore, the return to skill increases most in locations that choose to adopt PC most intensively. However, the relationship between skill supply and return to skill is weakly decreasing. After the arrival of the PC technology, the relationship between the supply of skill and the return to skill is given by

$$\ln \frac{W_H}{W_L} = \begin{cases} \ln \left[\frac{aH^{\rho-1}}{(1-a)L^{\rho-1}} \right] & \text{if } \frac{H}{L} \leq \phi^L \\ \ln \left[\frac{a(\phi^L)^{\rho-1}}{(1-a)} \right] = \ln \left[\frac{b(\phi^H)^{\rho-1}}{(1-b)} \right] & \text{if } \phi^L < \frac{H}{L} < \phi^H \\ \ln \left[\frac{bH^{\rho-1}}{(1-b)L^{\rho-1}} \right] & \text{if } \frac{H}{L} \geq \phi^H \end{cases} \quad (50)$$

where ϕ^H and ϕ^L are the critical values of the skill ratio such that if a location is characterized by $\frac{H}{L} < \phi^L$, then it continues to use the old technology. If $\frac{H}{L} > \phi^H$, then the location switches to the new technology. Equation (50) shows that when a firm keeps the old technology, the relationship between the skill ratio and the skill premium is negative, and the same remains true once the firm has fully switched to the new technology. However, when the firm is in transition between the old and new technologies, the slope is zero. This transition-and-adoption mechanism therefore does not reproduce the positive relationship between the skill ratio and the skill premium documented in Figure 3. Complementary papers such as Rubinton (2025) and Eckert, Ganapati and Walsh (2022) develop quantitative models studying the role of capital-skill complementarity in space.

D.2 Skill Accumulation over Time

The share of highly skilled workers has increased dramatically over the past fifty years. This secular rise in educational attainment could, in principle, alter the local skill composition and thus contribute to regional divergence independently of agglomeration or technology diffusion. In particular, if cities that initially had more college-educated workers subsequently experienced faster skill upgrading, the observed divergence could reflect differences in the pace of local skill accumulation rather than the interaction of SBTC and agglomeration emphasized in the model.

To compare the predictive content of initial skill composition and subsequent skill growth, I augment the β -convergence regression in Appendix Table D.3 with both the initial skill ratio, $(H/L)_{j,1980}$, and subsequent annualized skill growth, $\Delta \log(H/L)_{j,1980-2010}/30$. The dependent variable is annualized wage growth across MSAs between 1980 and 2010. The Shapley calculation allocates the increase in R^2 associated with adding these two controls, averaging across their possible entry orderings. The initial skill ratio receives the larger share of this incremental explanatory power. This comparison concerns the explanatory power of the two controls; it does not identify the separate causal effects of SBTC, agglomeration, or human-capital accumulation. I examine the role of aggregate skill upgrading in the calibrated-model counterfactual below.

Table D.4 reports a separate calibrated-model counterfactual that holds post-1980 aggregate skill shares fixed at their 1980 levels. Relative to the baseline, this exercise lowers the high-skill convergence change. The comparison measures the contribution of aggregate skill growth within the calibrated model.

The low-skill response moves in the same numerical direction as the high-skill response but has a different economic interpretation. In the model, the low-skill convergence change is $\Delta\beta^L = -0.36$ pp (-23.28%), whereas under No H/L Growth it is $\Delta\beta^L = -0.73$ pp (-47.48%). Both are negative, and fixing aggregate skill growth makes the low-skill change more negative, strengthening convergence rather than reversing its sign.

Panel B shows that secular migration responses are nearly unchanged: $\Delta Mig.^H$ changes from -0.61 pp in the model to -0.61 pp, while $\Delta Mig.^L$ changes from -0.37 pp to -0.38 pp. The skill-gap convergence moment moves from 0.37 pp to 0.27 pp. Together with Panel A, these outcomes show that fixing aggregate skill growth attenuates the high-skill reversal and strengthens the low-skill convergence change.

D.3 International Trade

International trade has been studied as a contributor to changes in relative labor demand and the national skill premium (Feenstra and Hanson 1999), and could therefore also affect regional β -convergence. Autor, Dorn and Hanson (2015) show that rising import penetration from China affected employment in more exposed commuting zones. I examine this possibility in two ways. First, I estimate the state-level β -convergence regression while controlling for import penetration from China as in Autor, Dorn and Hanson (2015); the estimated convergence coefficient changes little. Second, I evaluate the model-implied contribution of the post-1980 opening to the Rest of the World in the counterfactual below.

Counterfactual: Shutting down Trade with the Rest of the World To quantify the importance of the opening to international trade for the decline in β -convergence, I augment the domestic economy with a single “Rest of the World” (ROW) region. ROW is represented as a competitive supplier with sectoral producer prices p_d^{ROW} fixed exogenously and normalized to one. Fixing these prices makes ROW infinitely elastic at the normalized world prices. Iceberg trade costs $\tau_{d,i\text{ROW}}$ connect ROW to each U.S. city i and industry d . Given domestic producer prices $p_{dj,t}$, the Armington price index in city i , sector d , and year t is

$$P_{id,t} = \left[\sum_j \left(\tau_{d,ij} p_{dj,t} \right)^{1-\sigma_d} + s_t^{\text{ROW}} \left(\tau_{d,i\text{ROW}} p_d^{\text{ROW}} \right)^{1-\sigma_d} \right]^{\frac{1}{1-\sigma_d}}, \quad (51)$$

where $\tau_{d,ij}$ are domestic trade costs and σ_d is the sectoral elasticity of substitution across varieties. In the baseline I set $s_t^{\text{ROW}} = 0$ for $t < 1980$ and $s_t^{\text{ROW}} = 1$ for $t \geq 1980$, and I calibrate the wedges $\tau_{d,i\text{ROW}}$ so that in 1980 the model matches the observed U.S. import shares from ROW separately for goods (18%) and services (6%). All other parameters, including domestic trade costs $\tau_{d,ij}$, amenities, migration costs, and the paths of skill-biased technical change, are held fixed at their baseline values.

To quantify the contribution of the post-1980 opening to trade with ROW, I compute a counterfactual path in which ROW remains closed for the entire period. Formally, the counterfactual economy is identical to the baseline economy described in Section 3 and uses the same initial conditions and parameter sequences, except that I impose $s_t^{\text{ROW}} = 0$ for all t , so that the ROW term in (51) is never active. Equivalently, the counterfactual behaves as if $\tau_{d,i\text{ROW}} = +\infty$ at all dates. I then resolve the full dynamic spatial equilibrium from 1960 to 2010 under this trade regime.

The effect of the ROW opening is measured by comparing the baseline and counterfactual paths for the convergence statistics introduced in Section 2.2 (the β , β^H , and β^L coefficients), as well as the evolution of city-level wages, population, and secular migration. The counterfactual measures the model-implied contribution of the expansion of trade with the Rest of the World after 1980, holding skill-biased technical change, domestic trade costs, and aggregate skill supplies fixed.

Table D.6 indicates that shutting down trade openness to the Rest of the World attenuates, but does not eliminate, the baseline post-1980 skill-specific convergence reversal. In the baseline, the post-1980 changes are $\Delta\beta^H = 2.93$ pp (154.74%) and $\Delta\beta^L = -0.36$ pp (−23.28%), while under No ROW they are $\Delta\beta^H = 2.91$ pp (153.67%) and $\Delta\beta^L = -0.37$ pp (−23.67%). This interpretation remains consistent with the reduced-form evidence in Table D.5, where controlling for import penetration does not overturn the main convergence estimates.

D.4 The Importance of Local Skill-Biased Technical Change

This subsection studies whether the quantitative results in the baseline economy are sensitive to the level at which skill-biased technical change (SBTC) is measured and fed into the model. This addresses a natural concern: because local labor markets differ in their occupation and industry composition, a purely national SBTC shifter may mask meaningful heterogeneity in exposure to computerization across cities and sectors.

Starting from the baseline city-level SBTC series, I compare alternative constructions while holding all other parameters and initial conditions fixed and recomputing the equilibrium path and the convergence and migration objects reported in Table D.7.

Primitive inputs and notation. Let $L_{j\nu dt}^k$ denote employment of skill group $k \in \{H, L\}$ in occupation ν , industry d , and city j at time t . Let $\mathcal{O}_d$ denote the set of occupations observed in industry d . Finally, let R_ν denote a routine classification for occupation ν (based on the routine-task intensity (RTI) classification used in the paper), and $\mathbf{1}\{R_\nu = 1\}$ an indicator for routine-intensive occupations.

Four alternative SBTC constructions. I construct a sequence of skill-specific SBTC objects that differ only in the level of aggregation across cities and/or industries, starting from the main specification in (53). These are the SBTC blocks used in the robustness exercises underlying Table D.7.

(i) *City×industry SBTC (finest measure).* The most disaggregated object is the within- (j, d, t, k) routine-employment share:

$$\text{SBTC}_{jdt}^k = \frac{\sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k \mathbf{1}\{R_\nu = 1\}}{\sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k}, \quad k \in \{H, L\}. \quad (52)$$

By construction, (52) preserves (a) cross-city heterogeneity within the same industry and (b) cross-industry heterogeneity within the same city. This SBTC measure relates exposure to computerization to local industry×occupation structure.

(ii) *City-level SBTC (baseline “City” specification in Table D.7).* The city-level object collapses (52) across industries. Equivalently, it is the routine-employment share for each skill group within city j :

$$\text{SBTC}_{jt}^k = \frac{\sum_d \sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k \mathbf{1}\{R_\nu = 1\}}{\sum_d \sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k}, \quad k \in \{H, L\}. \quad (53)$$

This is the natural “local SBTC” measure: it allows SBTC to vary by city and skill over time, abstracting from the within-city industry decomposition.

(iii) *High- S_{city}^H stress test (City High- S^H in Table D.7).* To assess how sensitive the model's spatial predictions are to the intensity of highly skilled SBTC (as opposed to its cross-sectional allocation), I implement a stress test that scales up the city-level SBTC series for highly skilled workers in equation (53) by 15%:

$$\widetilde{\text{SBTC}}_{jt}^H = 1.15 \cdot \text{SBTC}_{jt}^H, \quad \widetilde{\text{SBTC}}_{jt}^L = \text{SBTC}_{jt}^L. \quad (54)$$

This counterfactual mechanically raises the skill premium component coming from the SBTC block, while leaving all other fundamentals unchanged, and therefore provides a direct way to gauge how much additional skill-biased innovation favoring highly skilled workers would be needed to materially alter the spatial convergence patterns.

(iv) *Industry-only SBTC (common across cities within industry).* The industry-level object removes cross-city heterogeneity within each industry by aggregating (52) across cities using industry-skill employment weights:

$$\text{SBTC}_{dt}^k = \frac{\sum_j \left(\sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k \right) \text{SBTC}_{jdt}^k}{\sum_j \sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k} = \frac{\sum_j \sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k \mathbf{1}\{R_\nu = 1\}}{\sum_j \sum_{\nu \in \mathcal{O}_d} L_{j\nu dt}^k}, \quad k \in \{H, L\}, \quad (55)$$

which isolates sectoral technological shifts that are common across cities. Empty city-industry cells receive zero weight automatically under this aggregation. Under this construction, any spatial heterogeneity in SBTC exposure arises only through differences in cities' industry composition (i.e., a pure industry shift-share channel).

(v) *National SBTC (common across cities and industries).* Finally, the national SBTC series is constructed as the (unweighted) cross-city average of the city-level object (53) in each year:

$$\text{SBTC}_t^k = \frac{1}{J} \sum_{j=1}^J \text{SBTC}_{jt}^k, \quad k \in \{H, L\}. \quad (56)$$

This is the most aggregated shifter and corresponds to imposing a single time series of SBTC for each skill group, obtained by averaging the city-level SBTC measure within year before feeding it into the model.

Implementation in the quantitative model. For each SBTC construction above, I feed the resulting shifters into the same estimated baseline economy. Concretely, I keep fixed: (i) the estimated agglomeration elasticities and diffusion parameter (Section 4 and Appendix Section C), (ii) the migration-cost and amenity objects obtained by inversion (Appendix Section C.4), (iii) the full trade block and the calibrated trade-cost matrix, and (iv) the initial

productivity distribution and all non-SBTC fundamentals.

The only change across specifications in Table D.7 is the level at which the SBTC shifter varies: by (j, t) in the City and High- S^H -City cases, by (j, d, t) in the City $\times$ Industry case, by (d, t) in the Industry-only case, and only by t in the National case. All equilibrium objects and the convergence coefficients reported in Table D.7 are then computed exactly as in the baseline counterfactual section (same regression specification, same population weights, same time windows).

Table D.7 summarizes how the model’s key spatial moments change when the SBTC shifter is constructed at different levels of aggregation. The top panel reports the raw early and late 20-year coefficients and their changes, while the bottom panel reports the associated allocation and sorting outcomes.

(1) City (baseline). The City specification uses the city-level exposure indices (j, t) defined in (53). The baseline model displays a post-1980 slowdown in convergence concentrated in high-skill wages: $\Delta\beta_{2010-1980}^H = 2.93$ pp. Measured relative to the pre-1980 magnitude, this corresponds to 154.74%. Less skilled wage convergence strengthens modestly, with a change of $\Delta\beta_{2010-1980}^L = -0.36$ pp (about -23.28% of the pre-1980 magnitude). Consistent with the mechanism emphasized in the main text, the bottom panel shows that this break is accompanied by sizable changes in average secular migration between 1960–1980 and 1990–2010 ($\Delta Mig.^H = -0.61$ pp and $\Delta Mig.^L = -0.37$ pp) and by divergence in the local skill ratio ($\Delta\beta_{80-10}^{HL} = 0.37$ pp).

(2) High- S^H City. The High- S^H City specification rescales the high-skill city series by 15% as in (54), increasing the intensity of high-skill SBTC while preserving its spatial ranking. The implied convergence and sorting moments are very close to the baseline City case (e.g., $\Delta\beta^H$ moves from 2.93 to 2.93 pp), indicating limited local sensitivity to this particular perturbation of the high-skill exposure index.

(3) City $\times$ Industry. The City $\times$ Industry specification allows SBTC to vary at the (j, d, t) level, so that cities can experience different SBTC intensity within the same industry. The resulting high-skill and low-skill changes are $\Delta\beta^H = 2.86$ pp and $\Delta\beta^L = -0.47$ pp, respectively; relative to their pre-1980 magnitudes, the corresponding changes are 154.71% and -31.01% .

(4) Industry-only. The Industry-only specification shuts down within-industry cross-city differences and lets SBTC vary only by industry and time (d, t) (see (55)). The resulting high-skill and low-skill changes are $\Delta\beta^H = 2.86$ pp and $\Delta\beta^L = -0.48$ pp; relative to their pre-1980 magnitudes, they are 150.38% and -33.16% . Relative to City $\times$ Industry, Industry-only changes the magnitudes of both skill-group responses while preserving their signs.

(5) National. Finally, the National specification imposes a purely national SBTC path

(see (56)). Even without cross-city variation in SBTC exposure, the model still generates a sizable slowdown in highly skilled wage convergence ($\Delta\beta^H = 2.88$ pp and $\Delta\beta^L = -0.46$ pp), because a common skill-biased shock interacts with heterogeneous local agglomeration forces, migration frictions, and diffusion. Relative to the City construction, the National specification reports $\Delta\beta^H = 2.88$ pp and $\Delta\beta^L = -0.46$ pp; the corresponding City values are 2.93 pp and -0.36 pp. The national specification changes the size of the convergence response differently across skill groups.

Quantitative summary (percent changes). A key takeaway from Table D.7 is that the *percent changes* vary across SBTC constructions. For high-skill wages, the City, High- S^H City, City $\times$ Industry, Industry, and National values are 154.74%, 154.76%, 154.71%, 150.38%, and 151.73%. The corresponding low-skill values are -23.28% , -23.23% , -31.01% , -33.16% , and -31.47% . The magnitudes differ across constructions, but the skill-specific sign pattern is unchanged: highly skilled convergence weakens while less skilled convergence strengthens.

D.5 Housing Regulation

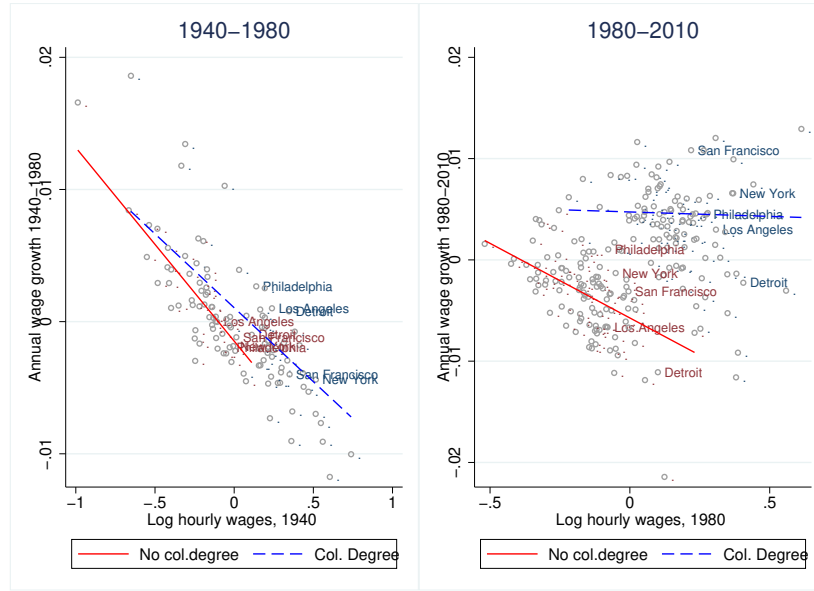
Ganong and Shoag (2017) provide an explanation based on housing prices that suggests that the U.S. states where housing prices increased the most are also the ones where migration declined. Hence, because migration increases convergence, the decline in migration to these areas, which are also the richest,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

I also examine the skill-specific convergence patterns for MSAs in states with low housing-supply elasticities. Figure D.1 shows that the asymmetry between highly skilled and less skilled wage convergence remains visible in this subgroup. This descriptive comparison does not identify the causal effect of housing regulation, but it indicates that the central skill-specific pattern is not confined to states with relatively elastic housing supply.

D.6 Structural Transformation Towards IT Service and Communication

Technological innovation might have caused a sectoral rather than a skill-biased effect. Such an effect would cause productivity increases in highly innovative industries such as IT service and communication. Therefore, cities with a higher concentration of creative industries benefit more from technological change. To control for the IT sector's importance, I estimate

Figure D.1: Regional β -convergence across MSAs before and after 1980 by Skill Group: Low-Housing-Elasticity States



Note: Figure D.1 shows two scatter plots of log wages by MSA in the initial year against annual average wage growth between the initial and final years by skill type (highly skilled and less skilled workers) in cities that are in states with low housing elasticities. Specifically, on the left-hand side (right-hand side), I plot the demeaned log wages in 1940 (1980) by MSA against the annual average growth of wages between 1940 (1980) and 1980 (2010) by skill type (highly skilled and less skilled workers). The line in each graph represents a weighted regression line from the bivariate regression.

β -convergence regressions with controls in wages between 1980 and 2010.² The results reported in Table D.8 are estimated at the MSA $\times$ IT $\times$ education-cell level. Column (1) already shows statistically significant divergence: cells with higher initial wages in 1980 experienced faster subsequent wage growth. Adding the IT-sector dummy in column (2) leaves this result intact, although the coefficient becomes somewhat smaller. In column (3), adding the college-degree dummy reverses the sign on initial wages, indicating that skill composition, rather than IT intensity per se, accounts for an important part of the raw divergence across cells. This evidence suggests that sectoral differences in technological intensity alone do not explain the regional convergence patterns. The framework developed above allows skill intensities to differ across the broad industries listed in Table C.1.

²I define the IT sector by looking at the codes of the IND1990 variable in the IPUMS data set and select industries that are more technology-oriented.

Table D.1: β -Convergence Rates Across States

Panel A: Convergence Rate						
	Population-weighted			Unweighted		
	(1)	(2)	(3)	(4)	(5)	(6)
	1940-1980	1980-2010	1980-2010-IV	1940-1980	1980-2010	1980-2010-IV
$\ln(wage^{1940})$	-1.788*** (-15.32)			-1.747*** (-18.59)		
$\ln(wage^{1980})$		-1.001** (-3.41)	-0.0926 (-0.24)		-0.557 (-1.54)	0.0315 (0.07)
Panel B: Convergence Rate by Skills: β^H and β^L						
	Population-weighted					
	1940-1980		1980-2010			
	No College	College	No College	College		
Panel B.1						
$\ln(wage^{1940})$	-1.823*** (-21.86)	-1.146*** (-7.58)				
$\ln(wage^{1980})$			-1.542*** (-8.43)	-1.125*** (-3.84)		
	Unweighted					
	1940-1980		1980-2010			
	No College	College	No College	College		
Panel B.2						
$\ln(wage^{1940})$	-1.815*** (-23.59)	-1.068*** (-8.48)				
$\ln(wage^{1980})$			-1.337*** (-5.27)	-0.704** (-2.25)		
N	49	49	51	51		

Note: Table D.1 reports the estimates of the β -convergence plotted in Figures 1 and 2 across states rather than for MSAs as in the figures. Panel A reports aggregate state-level convergence. Columns (1)–(3) are population-weighted estimates: column (1) is the 1940–1980 OLS specification, column (2) is the 1980–2010 OLS specification, and column (3) is the 1980–2010 IV specification using wages in 1970 as the instrument for initial wages in 1980. Columns (4)–(6) report the corresponding unweighted specifications. Panel B reports skill-specific state-level convergence. Panel B.1 reports the population-weighted estimates, with columns (1)–(2) corresponding to No College and College in 1940–1980 and columns (3)–(4) corresponding to No College and College in 1980–2010. Panel B.2 reports the same four regressions without population weights. All standard errors are robust, and t -statistics are in parentheses. In Panel A, ***, **, and * denote statistical significance at the 0.001, 0.01, and 0.05 levels, respectively. In both subpanels of Panel B, ***, **, *, and + denote significance at the 0.01, 0.05, 0.10, and 0.15 levels, respectively. The dependent variable in each regression is the annual average wage growth between the initial and final years reported at the top. All reported β -convergence coefficients are in annual percentage points (%).

Table D.2: The Impact of Physical Capital Transition Dynamics

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
With Capital	-0.67	-2.40	-1.73 pp (-257.50%)	-0.47	-0.26	0.21 pp (44.56%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
With Capital	-0.61 pp		-0.37 pp	0.37 pp		0.23

Note: Table D.2 compares Model with the With Capital specification (location-level installed capital and transition dynamics). Panel A reports the raw 1960–1980 and 1990–2010 β^H and β^L coefficients and their changes. Every regression uses endpoint-period skill-employment weights, with no temporal smoothing. The capital block operates from 1960, so its early coefficients need not equal Model. Changes are in percentage points. Panel B reports secular migration, skill-ratio convergence, and sorting outcomes.

Table D.3: Skill Composition and Wage Convergence

	(1)	(2)	(3)	(4)	(5)
$\ln(wage^{1940})$	-1.266*** (0.102)				
$\ln(wage^{1980})$		-0.122 (0.494)	-1.357*** (0.332)	-0.302 (0.505)	-1.686*** (0.294)
Initial H/L, 1980			3.165*** (0.395)		3.379*** (0.340)
Annual H/L Growth, 1980–2010				18.764*** (4.658)	23.452*** (3.022)
Observations	132	247	239	239	239
R-squared	0.638	0.001	0.424	0.098	0.575

Note: Table D.3 reports the estimates of β -convergence controlling for the initial level of skill ratio in 1980 and for annual skill ratio growth between 1980 and 2010. Column (1) reports the β -convergence estimate without controls for 1940–1980 using 1940 population weights. Columns (2)–(5) report the β -convergence for the 1980–2010 period with different sets of controls using 1980 population weights. All reported β -convergence coefficients are in annual percentage points (%). The table also reports the corresponding R-squared values used in the Shapley decomposition discussed in the text.

Table D.4: Impact of Increase in Skill Ratio on Regional Convergence

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
No H/L Growth	-1.90	0.60	2.50 pp (131.81%)	-1.55	-2.28	-0.73 pp (-47.48%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
No H/L Growth	-0.61 pp		-0.38 pp	0.27 pp		0.08

Note: Table D.4 compares Model with No H/L Growth, which fixes national skill totals at their 1980 levels from 1980 onward. Panel A reports raw 1960–1980 and 1990–2010 convergence coefficients and their changes using endpoint-period skill-employment weights, with no temporal smoothing. The pre-1980 path is unchanged. Panel B reports secular migration, skill-ratio convergence, and sorting outcomes.

Table D.5: β -Convergence and Import Penetration from China

	(1)	(2)	(3)	(4)
$\ln(wage^{1980})$	-0.840*** (-2.74)	-0.760** (-2.58)	-0.604** (-2.26)	-0.539** (-2.12)
Import Penetr.		-0.0320 (-1.34)		-0.0373 (-1.42)
Pop. Weight	Yes	Yes		
Observation	49	48	49	48
R square	0.26	0.27	0.14	0.16

Note: Table D.5 shows the results for a β -convergence regression of wage growth between 1980 and 2010 on the initial wage in 1980, conditioning on the import penetration from China by the state. All reported β -convergence coefficients are in annual percentage points (%).

Table D.6: Impact of Trade with the Rest of the World on Regional Convergence

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
No ROW	-1.90	1.02	2.91 pp (153.67%)	-1.55	-1.92	-0.37 pp (-23.67%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
No ROW	-0.61 pp		-0.37 pp	0.36 pp		0.19

Note: Table D.6 compares Model with No ROW, which shuts external trade for the entire horizon. Panel A reports raw 1960–1980 and 1990–2010 convergence coefficients and their changes using endpoint-period skill-employment weights, with no temporal smoothing. Because the intervention operates from 1960, the early coefficients may differ. Panel B reports secular migration, skill-ratio convergence, and sorting outcomes.

Table D.7: Alternative Definition of SBTC

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
SBTC _{city} ^{high S^H (1980+)}	-1.90	1.04	2.93 pp (154.76%)	-1.55	-1.91	-0.36 pp (-23.23%)
SBTC _{city×industry}	-1.85	1.01	2.86 pp (154.71%)	-1.51	-1.98	-0.47 pp (-31.01%)
SBTC _{industry}	-1.90	0.96	2.86 pp (150.38%)	-1.46	-1.95	-0.48 pp (-33.16%)
SBTC _{national}	-1.90	0.98	2.88 pp (151.73%)	-1.46	-1.93	-0.46 pp (-31.47%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
SBTC _{city} ^{high S^H (1980+)}	-0.61 pp		-0.37 pp	0.37 pp		0.20
SBTC _{city×industry}	-0.61 pp		-0.37 pp	0.34 pp		0.15
SBTC _{industry}	-0.61 pp		-0.37 pp	0.39 pp		0.17
SBTC _{national}	-0.61 pp		-0.37 pp	0.38 pp		0.17

Note: Table D.7 reports Model under five SBTC definitions: City, City×Industry, City High- S^H , Industry, and National. Panel A reports raw 1960–1980 and 1990–2010 convergence coefficients and their changes using endpoint-period skill-employment weights, with no temporal smoothing. Panel B reports secular migration, skill-ratio convergence, and sorting outcomes.

Table D.8: Convergence Rates and IT

	(1) 1980–2010	(2) 1980–2010	(3) 1980–2010
$\ln(wage^{1980})$	0.80*** (2.95)	0.64** (2.52)	-1.35*** (-3.00)
IT		0.40*** (10.94)	0.56*** (17.14)
College Degree			1.13*** (8.85)
Pop. Weight	Yes	Yes	Yes
Observations	985	985	985
R-squared	0.07	0.15	0.43

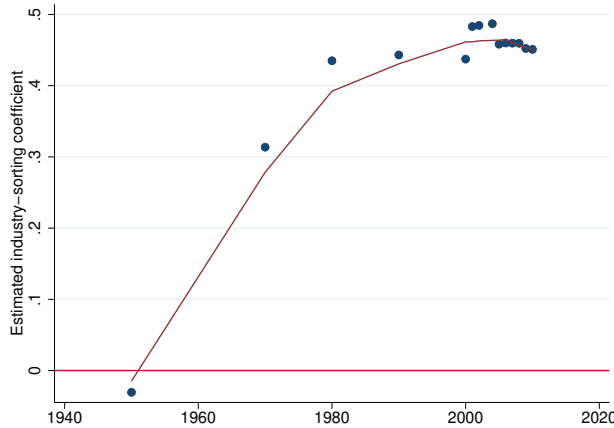
Note: Table D.8 reports β -convergence regressions for wage growth between 1980 and 2010 estimated at the MSA×IT×education-cell level. The dependent variable is the annualized growth rate in wages for each cell, multiplied by 100. Column (1) regresses wage growth on initial log wages only. Column (2) adds the IT-sector dummy. Column (3) adds both the IT-sector dummy and the college-degree dummy. Regressions are weighted by the 1980 cell population and standard errors are clustered at the MSA level, which is why the number of observations exceeds the number of MSAs.

In addition to sectoral innovation shifts, changes in firms’ relocation decisions over time could affect regional β -convergence. Identifying such changes requires firm-level panel data. Faberman and Freedman (2016) use longitudinal establishment data for the U.S. during 1992–1997 and do not find that spillovers are important for firms’ decisions to locate in urban rather than nonurban areas. Comparable firm-level data extending back to 1940 are unavailable. I therefore use publicly available industry-level data to test whether more skill-intensive industries have become increasingly concentrated in more skill-intensive cities over time. This exercise measures alignment between city and industry skill composition; it does not identify firm relocation. I estimate the following interaction regression:

$$\text{Skill concentration}_{djt} = \alpha + \sum_{\tau=1950}^T \beta_{\tau} \mathbf{1}(t = \tau) \left(\frac{H_{jt}}{L_{jt}} \right) + \phi_t + \phi_j + \epsilon_{djt} \quad (57)$$

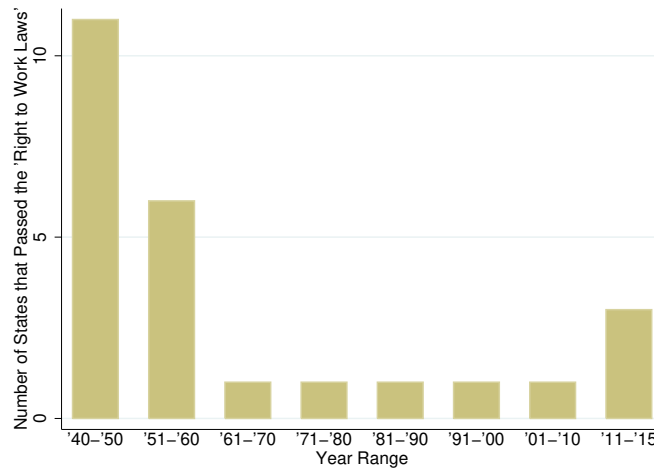
where d is the industry, j is the MSA, t is calendar time, and τ indexes the year-specific interaction coefficients. The ϕ_t are time fixed effects, and ϕ_j are MSA fixed effects. I build the measure of “Skill concentration” by calculating the ratio between the number of skilled workers and the total number of workers in industry d in location j at time t . Figure D.2 plots the estimated coefficients β_{τ} over time. The figure shows that industry skill intensity becomes more strongly associated with city skill intensity. This is evidence of increasing alignment between the skill composition of industries and cities, but not direct evidence of firm sorting.

Figure D.2: Industry Sorting over Time



Note: Figure D.2 plots over time the estimated coefficient on local skill concentration in the industry-sorting regression of equation (57). The sample combines decennial Census observations through 2000 with annual ACS observations for 2001, 2002, and 2004–2010; 2003 is omitted in the underlying regression sample. The vertical axis therefore reports the estimated industry-sorting coefficient.

Figure D.3: Right-to-Work Laws



Note: Figure D.3 reports a histogram of the number of states that adopted right-to-work laws by decade, beginning in the 1940s.

D.7 Right-to-Work Laws and Unions

In the Southern and Western U.S., 26 states have passed *Right-to-Work Laws* since 1940. These laws permit workers to work without having to join a union. They could affect the spatial distribution of less skilled wages, and [Holmes \(1998\)](#) shows that state policies influence industry location. Figure D.3 shows that most adopting states passed these laws in the 1940s and 1950s, before the secular decline in regional β -convergence. Union membership has also fallen substantially in the United States and could have affected regional convergence directly. I use CPS data on union membership aggregated to the state level beginning in 1990. Table D.9 reports state-level β -convergence regressions for 1990–2010 with and without this control. The estimated convergence coefficient changes little and is slightly less negative when the state union-membership measure is included.

Table D.9: β -Convergence and Unions

	(1)	(2)	(3)	(4)
$\ln(wage^{1990})$	-0.840*** (-2.76)	-0.723** (-2.31)	-0.604** (-2.28)	-0.454 (-1.63)
Union Presence		-4.131** (-2.27)		-4.275** (-2.28)
Pop. Weight	Yes	Yes		
Observation	147	147	147	147
R square	0.26	0.30	0.14	0.19

Note: Table D.9 reports estimates from regressions of annual wage growth between 1990 and 2010 on log initial wages in 1990, with and without the state-level union-membership control. All reported β -convergence coefficients are in annual percentage points (%).

D.8 Evidence of Spatial Technology Diffusion

This subsection presents evidence on spatial technology diffusion in two complementary steps. I first document the empirical geographic pattern of convergence in the data. I then show that removing diffusion over the full horizon materially changes both the pre-1980 and post-1980 paths.

D.8.1 Empirical Evidence

Introducing a persistent productivity process with spatial diffusion directly embeds regional convergence into the model, consistent with the mechanisms emphasized by Barro and Sala-i Martin (1997), Caselli and Coleman (2001), Desmet, Nagy and Rossi-Hansberg (2018), and Eckert and Peters (2022). Unlike frameworks that focus on steady-state comparisons, a diffusion-based formulation naturally captures dynamics in which productivity gradually spreads from more to less advanced locations. This approach follows the intuition in Barro and Sala-i Martin (1997), who show that a neoclassical model with frictions to capital mobility can reproduce empirical convergence rates across countries and within the United States. Similarly, Caselli and Coleman (2001) develop a dynamic model in which total factor productivity (TFP) grows faster in agriculture, the cost of human-capital accumulation declines over time, and farm goods are a necessary input, all of which generate convergence through adjustment dynamics. Relatedly, Caliendo (2011) and Bajona and Kehoe (2010) demonstrate that convergence can also arise endogenously in dynamic Heckscher–Ohlin frameworks. In this paper, the convergence mechanism generated by idea diffusion is conceptually closest to a setting with declining barriers to human- or physical-capital mobility.

To provide supporting evidence for spatial diffusion in the U.S. data, I estimate two comple-

mentary reduced-form exercises summarized in Table D.10. Panel A reports $\beta_{\tau,t}$, $\beta_{\tau,t}^{\text{within}}$ (with state fixed effects), and their ratio $\beta_{\tau,t}^{\text{within}}/\beta_{\tau,t}$. This ratio compares convergence coefficients with and without state fixed effects; it is not a within/between variance decomposition. In 1940–1960 the ratio is about 0.98, so the coefficient estimated with state fixed effects is close to the corresponding coefficient without them; the ratio then declines over time. Panel B estimates an all-city interaction specification in which wage growth is regressed on distance to high-wage cities, an indicator for low-wage cities, and their interaction. This specification keeps the full sample while separately identifying the distance gradient for low-wage locations. In the preferred specification, I focus on decade-by-decade windows and report no-region-FE estimates with objective outlier trimming (Cook’s distance and leverage). One possible explanation for the weak positive sign in 1940–1950 is that wartime and immediate postwar reallocation affected local production and migration patterns in ways not captured by a peacetime distance-to-frontier diffusion channel. Panel B shows a negative distance gradient for low-wage locations in 1950–1960. The gradients in 1960–1970 and 1970–1980 are close to zero and imprecisely estimated, providing limited evidence of diffusion after 1960. The within-state diagnostic in Panel A, the skill-specific persistence estimates, and the model comparison that shuts down diffusion provide complementary evidence on this mechanism.

The reduced-form evidence above isolates the geographic footprint of convergence in the data. I now connect this evidence to the model by comparing the baseline economy with a version that removes technology diffusion.

D.8.2 Quantitative Comparison Without Technology Diffusion

In this whole-horizon exercise, I set both persistence operators to one while holding the realized innovations fixed. Table D.11 compares Model and No Diffusion. Removing diffusion from 1960 changes convergence both before and after 1980, as well as wage dispersion and sorting. It produces strong early convergence for highly skilled wages and substantial divergence for less skilled wages, followed by divergence for both groups in the late window. Among the other reported outcomes, the largest difference from the baseline is the 2010 variance of less skilled wages.

Overall, removing the global-mean convergence term while holding the realized innovations fixed does not reproduce the observed joint H/L pattern and creates large dispersion differences.

Table D.11 shows early convergence for highly skilled wages and divergence for less skilled wages, followed by divergence for both groups. Wage dispersion and sorting also change substantially. The results show how diffusion shapes convergence and wage dispersion over the full simulation period. Because the intervention begins in 1960, its effects combine changes before and after 1980.

Table D.10: Evidence on Spatial Technology Diffusion

Panel A. Within-state component of convergence

	40-60	50-70	60-80	80-10
β	-2.44	-2.18	-2.15	-0.16
β_{within}	-2.38	-1.80	-1.12	0.06
β_{within}/β	0.98	0.83	0.52	-0.41

Panel B. All-city proximity gradient with low-wage interaction

	40-50	50-60	60-70	70-80	80-90	90-00	00-10
$\log w_{j,t_0}$	-3.14*** (0.62)	-0.90 (0.81)	-0.83 (0.78)	-2.94*** (0.80)	0.58 (1.43)	0.52 (0.74)	0.27 (0.46)
$\ln d_{j,t_0}^H$	0.18*** (0.06)	-0.03 (0.08)	-0.01 (0.07)	0.04 (0.07)	-0.06 (0.12)	-0.02 (0.09)	-0.08 (0.07)
$\mathbf{1}\{Low\} \times \ln d_{j,t_0}^H$	0.21 (0.19)	-0.24* (0.14)	0.06 (0.13)	0.05 (0.12)	0.22 (0.15)	0.19 (0.12)	0.08 (0.14)
$\beta_{Low}^{dist} = \beta_0 + \beta_1$	0.40** (0.18)	-0.28** (0.12)	0.05 (0.11)	0.09 (0.11)	0.16* (0.09)	0.17** (0.07)	0.00 (0.12)
Obs.	126	131	99	102	206	216	256
Region FE	No	No	No	No	No	No	No

Note: Panel A reports overall and within-state convergence slopes over rolling windows (40–60, 50–70, 60–80, 80–10), together with the ratio β^{within}/β ; this ratio measures how much of the overall slope remains after adding state fixed effects, not a formal within/between variance decomposition. All Panel A coefficients are reported in annual percentage points (%), except the ratio β^{within}/β . Panel B uses all cities and estimates decade-specific regressions of annual wage growth on initial log wage and distance to the nearest high-wage city, with an interaction between distance and a low-wage indicator: $gr_j = \alpha + \beta_0 \ln d_{j,t_0}^H + \beta_1 \mathbf{1}\{Low_j\} \times \ln d_{j,t_0}^H + \beta_2 \log w_{j,t_0} + \varepsilon_j$. Low-wage cities are the bottom 33rd percentile of the initial wage distribution; high-wage cities are the top 15th percentile (within each decade pair). The coefficient on $\ln d_{j,t_0}^H$ is the distance gradient for non-low-wage cities; β_1 is the additional distance gradient for low-wage cities. The row $\beta_{Low}^{dist} = \beta_0 + \beta_1$ reports the implied total distance gradient for low-wage cities (from a linear combination test). In Panel B, coefficients and standard errors are reported in annual percentage points (%). In Panel B, direct distance-gradient evidence is strongest in 1950–1960; later pre-1980 decade estimates are close to zero and imprecise. The ***, **, and * represent statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table D.11: A Comparison with No Technological Diffusion

	β_{1980}^H	β_{2010}^H	$\Delta\beta^H$	β_{1980}^L	β_{2010}^L	$\Delta\beta^L$
Model	-1.90	1.04	2.93 pp (154.74%)	-1.55	-1.91	-0.36 pp (-23.28%)
No Diffusion	-5.21	3.25	8.46 pp (162.28%)	9.72	1.83	-7.89 pp (-81.18%)
	$\Delta Mig.^H$ (p.p.)		$\Delta Mig.^L$ (p.p.)	$\Delta\beta^{HL}$ (p.p.)		$\Delta\delta$
Model	-0.61 pp		-0.37 pp	0.37 pp		0.20
No Diffusion	-1.74 pp		-0.48 pp	8.31 pp		0.09

Note: Table D.11 compares Model with No Diffusion, which sets both skill-specific persistence operators to one while holding the realized innovations fixed. Panel A reports raw 1960–1980 and 1990–2010 convergence coefficients and their changes using endpoint-period skill-employment weights, with no temporal smoothing. The intervention operates from 1960, so its early coefficients may differ. Panel B reports secular migration, skill-ratio convergence, and sorting outcomes.

E Analytical Illustration: Skill Bias, Agglomeration, and Wage Responses

This appendix presents a comparative-static illustration of how skill composition and city size shape wage responses to a joint change in national skill bias and local productivity shifters. Holding city populations fixed, the propositions characterize wage growth and skill-premium changes relative to a common initial-productivity benchmark. The composition and size shifters capture the local agglomeration channels in Section 3. The full quantitative model incorporates migration, endogenous sorting, housing, trade, and technology diffusion to study how these outcomes evolve over time.

E.1 Baseline Two–Skill, Two–Period Model

Environment. Cities $j \in \{1, \dots, J\}$ are populated by fixed masses $H_j, L_j > 0$ of highly skilled and less skilled workers. Let the skill ratio be $r_j \equiv H_j/L_j$. Output is produced under CES technology

$$Y_{jt} = \left(A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho \right)^{1/\rho}, \quad 0 < \rho < 1, \quad \sigma \equiv \frac{1}{1-\rho} > 1.$$

Firms are perfectly competitive and pay factor marginal products:

$$w_{Hjt} = A_{Hjt} H_j^{\rho-1} (A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho)^{\frac{1}{\rho}-1}, \quad w_{Ljt} = A_{Ljt} L_j^{\rho-1} (A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho)^{\frac{1}{\rho}-1}.$$

At $t = 0$, let $A_{Hj0} = A_{Lj0} = 1$. At $t = 1$, introduce a common skill-bias factor λ together with composition-dependent productivity shifters:

$$A_{Hj1} = \lambda r_j^{\gamma_H}, \quad A_{Lj1} = r_j^{\gamma_L}, \quad \lambda > 1, \quad \gamma_H > \gamma_L \geq 0.$$

Define $\pi_{jt} \equiv \log w_{Hjt} - \log w_{Ljt}$ as the log skill premium.

Proposition 1 (Skill premium and its change). *For any city j ,*

$$\pi_{jt} = \log\left(\frac{A_{Hjt}}{A_{Ljt}}\right) + (\rho - 1) \log r_j.$$

Hence, the change in the skill premium between $t = 0$ and $t = 1$ is

$$\Delta\pi_j = \pi_{j1} - \pi_{j0} = \log \lambda + (\gamma_H - \gamma_L) \log r_j,$$

which is strictly increasing in r_j when $\gamma_H > \gamma_L$.

Proof. Starting from the marginal products above, $\frac{w_{Hjt}}{w_{Ljt}} = \frac{A_{Hjt}}{A_{Ljt}} \left(\frac{H_j}{L_j}\right)^{\rho-1}$. Taking logs yields $\pi_{jt} = \log(A_{Hjt}/A_{Ljt}) + (\rho - 1) \log r_j$. At $t = 0$, $A_{Hj0} = A_{Lj0} = 1$; at $t = 1$, $A_{Hj1}/A_{Lj1} = \lambda r_j^{\gamma_H - \gamma_L}$. Subtracting gives the result. Monotonicity follows since $\partial\Delta\pi_j/\partial r_j = (\gamma_H - \gamma_L)/r_j > 0$ for $\gamma_H > \gamma_L$. $\square$

Economic intuition. The common skill-bias factor contributes $\log \lambda$ to every city's premium change. Introducing the composition-dependent shifters contributes $(\gamma_H - \gamma_L) \log r_j$, which accounts for the cross-city gradient.

Proposition 2 (Wage growth of each skill). *Let $\theta(r) \equiv r^\rho/(1 + r^\rho)$ denote the CES high-skill share term under the baseline normalization $A_{Hj0} = A_{Lj0} = 1$. Then:*

$$\begin{aligned} \Delta \log w_{Hj} &= \log \lambda + \gamma_H \log r_j + \frac{1 - \rho}{\rho} \log\left(\lambda r_j^{\gamma_H} \theta(r_j) + r_j^{\gamma_L} [1 - \theta(r_j)]\right), \\ \Delta \log w_{Lj} &= \gamma_L \log r_j + \frac{1 - \rho}{\rho} \log\left(\lambda r_j^{\gamma_H} \theta(r_j) + r_j^{\gamma_L} [1 - \theta(r_j)]\right). \end{aligned}$$

Moreover, on any domain where $\lambda r_j^{\gamma_H - \gamma_L} \geq 1$, both $\Delta \log w_{Hj}$ and $\Delta \log w_{Lj}$ are strictly increasing in r_j .

Proof. With $Y_{jt} = (A_{Hjt}H_j^\rho + A_{Ljt}L_j^\rho)^{1/\rho}$ and competitive factor markets,

$$w_{Hjt} = \frac{\partial Y_{jt}}{\partial H_j} = A_{Hjt}H_j^{\rho-1}(A_{Hjt}H_j^\rho + A_{Ljt}L_j^\rho)^{\frac{1}{\rho}-1}, \quad w_{Ljt} = A_{Ljt}L_j^{\rho-1}(A_{Hjt}H_j^\rho + A_{Ljt}L_j^\rho)^{\frac{1}{\rho}-1}.$$

Let $r_j \equiv H_j/L_j$ and normalize $L_j = 1$ (so $H_j = r_j$). At $t = 0$, $A_{Hj0} = A_{Lj0} = 1$ so

$$w_{Hj0} = r_j^{\rho-1}(1 + r_j^\rho)^{\frac{1}{\rho}-1}, \quad w_{Lj0} = (1 + r_j^\rho)^{\frac{1}{\rho}-1}.$$

At $t = 1$, $A_{Hj1} = \lambda r_j^{\gamma_H}$ and $A_{Lj1} = r_j^{\gamma_L}$, hence

$$w_{Hj1} = \lambda r_j^{\gamma_H} r_j^{\rho-1} (\lambda r_j^{\gamma_H+\rho} + r_j^{\gamma_L})^{\frac{1}{\rho}-1}, \quad w_{Lj1} = r_j^{\gamma_L} (\lambda r_j^{\gamma_H+\rho} + r_j^{\gamma_L})^{\frac{1}{\rho}-1}.$$

Taking log differences yields

$$\Delta \log w_{Hj} = \log \lambda + \gamma_H \log r_j + \frac{1-\rho}{\rho} \log F(r_j), \quad \Delta \log w_{Lj} = \gamma_L \log r_j + \frac{1-\rho}{\rho} \log F(r_j),$$

where

$$\theta(r) \equiv \frac{r^\rho}{1 + r^\rho}, \quad F(r) \equiv \lambda r^{\gamma_H} \theta(r) + r^{\gamma_L} (1 - \theta(r)) = \frac{\lambda r^{\gamma_H+\rho} + r^{\gamma_L}}{1 + r^\rho}.$$

Lemma 2 (Properties of θ). *Let $\theta(r) = \frac{r^\rho}{1+r^\rho}$ with $r > 0$ and $\rho \in (0, 1)$. Then*

$$\theta'(r) = \frac{\rho r^{\rho-1}}{(1 + r^\rho)^2} > 0, \quad (1 - \theta(r))' = -\theta'(r), \quad \frac{\partial \theta(r)}{\partial \log r} = r \theta'(r) = \rho \theta(r) (1 - \theta(r)).$$

Differentiate $F(r)$, using $(1 - \theta(r))' = -\theta'(r)$:

$$F'(r) = \lambda \gamma_H r^{\gamma_H-1} \theta(r) + \gamma_L r^{\gamma_L-1} (1 - \theta(r)) + \theta'(r) (\lambda r^{\gamma_H} - r^{\gamma_L}).$$

Since $\theta'(r) = \rho r^{\rho-1} (1 + r^\rho)^{-2} > 0$, $\theta(r) \in (0, 1)$, and $1 - \theta(r) \in (0, 1)$, the first two terms are nonnegative whenever $\gamma_H, \gamma_L \geq 0$. Moreover, on any domain where $\lambda r^{\gamma_H-\gamma_L} \geq 1$ (equivalently $\lambda r^{\gamma_H} \geq r^{\gamma_L}$), the third term is also nonnegative. Hence $F'(r) \geq 0$ on that domain, implying $\log F(r)$ is weakly increasing in r .

Therefore, on any domain where $\gamma_H, \gamma_L \geq 0$ and $\lambda r^{\gamma_H-\gamma_L} \geq 1$, both $\Delta \log w_{Hj}$ and $\Delta \log w_{Lj}$ are weakly increasing in r_j (strictly increasing if at least one inequality is strict). $\square$

Economic intuition. Wage growth for both groups rises with skill abundance because the CES aggregator places greater weight on the abundant factor and the high-skill technology grows faster.

Proposition 3 (Average of wage growths). *Let $\mu_H(r) = \frac{r}{1+r}$ and $\mu_L(r) = \frac{1}{1+r}$, and define $\Delta \log w_j = \mu_H(r_j) \Delta \log w_{Hj} + \mu_L(r_j) \Delta \log w_{Lj}$. If $\gamma_H > \gamma_L$ and the monotonicity conditions of Proposition 2 hold on the relevant support (so that $\frac{d}{dr} \Delta \log w_{Hj} \geq 0$ and $\frac{d}{dr} \Delta \log w_{Lj} \geq 0$), then $d \Delta \log w_j / dr_j > 0$.*

Proof. Recall $\Delta \log w_j = \mu_H(r_j) \Delta \log w_{Hj} + \mu_L(r_j) \Delta \log w_{Lj}$, where $\mu_H(r) = \frac{r}{1+r}$ and $\mu_L(r) = \frac{1}{1+r}$. Differentiating with respect to r_j yields

$$\frac{d}{dr} \Delta \log w_j = \mu'_H(r_j) (\Delta \log w_{Hj} - \Delta \log w_{Lj}) + \mu_H(r_j) \frac{d}{dr} \Delta \log w_{Hj} + \mu_L(r_j) \frac{d}{dr} \Delta \log w_{Lj}.$$

Since $\mu'_H(r) = \frac{1}{(1+r)^2} > 0$ and $\Delta \log w_{Hj} - \Delta \log w_{Lj} = \log \lambda + (\gamma_H - \gamma_L) \log r_j$, the first term is nonnegative on any domain where $\lambda r_j^{\gamma_H - \gamma_L} \geq 1$ (and strictly positive when the inequality is strict). Under the maintained assumptions (the same conditions invoked in Proposition 2) we also have $\frac{d}{dr} \Delta \log w_{Hj} \geq 0$ and $\frac{d}{dr} \Delta \log w_{Lj} \geq 0$. Therefore $\frac{d}{dr} \Delta \log w_j > 0$ on the stated domain. $\square$

Remark: initial average wages and β -patterns. In the baseline normalization $A_{Hj0} = A_{Lj0} = 1$, and setting $L_j = 1$ so that $H_j = r_j$, output is $Y_{j0} = (1 + r_j^\rho)^{1/\rho}$. The average wage (output per worker) is

$$w_{j0} = \frac{Y_{j0}}{1 + r_j} = \frac{(1 + r_j^\rho)^{1/\rho}}{1 + r_j}, \quad 0 < \rho < 1.$$

A direct derivative gives

$$\frac{d}{dr_j} \log w_{j0} = \frac{r_j^{\rho-1} - 1}{(1 + r_j^\rho)(1 + r_j)},$$

so w_{j0} is increasing in r_j for $r_j < 1$ and decreasing for $r_j > 1$. Consequently, the sign of a cross-sectional β -regression of average wage growth $\Delta \log w_j$ on initial average wages $\log w_{j0}$ is not pinned down mechanically by r_j alone; it depends on the joint distribution of (r_j, n_j) and other city fundamentals. The robust implication of the model is instead that $\Delta \log w_j$ (and, more sharply, the skill-premium change and high-skill wage growth) rises with initial skill abundance, as shown in Propositions 1–2 and 4–5.

Economic intuition. More skill-abundant cities have faster wage growth on the stated domain under the joint change in national skill bias and composition-dependent productivity.

Corollary 2 (Elasticity of substitution and the sensitivity of the skill premium). *Let $\sigma \equiv \frac{1}{1-\rho}$ denote the elasticity of substitution in the CES aggregator. Then, holding the technology ratio A_{Hjt}/A_{Ljt} fixed, the partial sensitivity of the (log) skill premium to relative skill supplies satisfies*

$$\frac{\partial \pi_{jt}}{\partial \log r_j} = \rho - 1 = -\frac{1}{\sigma}.$$

Hence, as σ increases, the magnitude of the dependence of the skill premium on r_j decreases.

Proof. From Proposition 1, $\pi_{jt} = \log(A_{Hjt}/A_{Ljt}) + (\rho - 1) \log r_j$. Taking the partial derivative with respect to $\log r_j$ while holding A_{Hjt}/A_{Ljt} fixed yields the result, and $\rho - 1 = -1/\sigma$. $\square$

E.2 Extension with Population Agglomeration Shifters

Keep the common initial benchmark $A_{Hj0} = A_{Lj0} = 1$ and let the date-one productivity shifters also depend on the fixed city population $n_j = H_j + L_j$:

$$A_{Hj1} = \lambda r_j^{\gamma_H} n_j^{\phi_H}, \quad A_{Lj1} = r_j^{\gamma_L} n_j^{\phi_L}, \quad \phi_H, \phi_L \geq 0.$$

Proposition 4 (Skill premium with agglomeration). *The change in the skill premium is*

$$\Delta \pi_j^A = \log \lambda + (\gamma_H - \gamma_L) \log r_j + (\phi_H - \phi_L) \log n_j,$$

which is strictly increasing in r_j if $\gamma_H > \gamma_L$ and in n_j if $\phi_H > \phi_L$.

Proof. Starting from the marginal products with agglomeration, $\frac{w_{Hj1}^A}{w_{Lj1}^A} = \frac{A_{Hj1}}{A_{Lj1}} \left(\frac{H_j}{L_j} \right)^{\rho-1}$. Taking logs yields $\pi_{j1}^A = \log \left(\frac{A_{Hj1}}{A_{Lj1}} \right) + (\rho - 1) \log r_j$. At $t = 0$, $A_{Hj0} = A_{Lj0} = 1$. At $t = 1$, $A_{Hj1} = \lambda r_j^{\gamma_H} n_j^{\phi_H}$ and $A_{Lj1} = r_j^{\gamma_L} n_j^{\phi_L}$, so

$$\log \left(\frac{A_{Hj1}}{A_{Lj1}} \right) = \log \lambda + (\gamma_H - \gamma_L) \log r_j + (\phi_H - \phi_L) \log n_j.$$

Subtracting the $t = 0$ expression gives the result. Monotonicity follows immediately from the coefficients on $\log r_j$ and $\log n_j$. $\square$

Proposition 5 (Wage growth of each skill with agglomeration). *Let $\theta(r) \equiv r^\rho / (1 + r^\rho)$ denote the CES high-skill share term under the baseline normalization $A_{Hj0} = A_{Lj0} = 1$, and define*

$$F^A(r, n) \equiv \lambda r^{\gamma_H} n^{\phi_H} \theta(r) + r^{\gamma_L} n^{\phi_L} (1 - \theta(r)).$$

Then:

$$\begin{aligned} \Delta \log w_{Hj}^A &= \log \lambda + \gamma_H \log r_j + \phi_H \log n_j + \frac{1 - \rho}{\rho} \log F^A(r_j, n_j), \\ \Delta \log w_{Lj}^A &= \gamma_L \log r_j + \phi_L \log n_j + \frac{1 - \rho}{\rho} \log F^A(r_j, n_j). \end{aligned}$$

Moreover, on any domain where $\gamma_H, \gamma_L, \phi_H, \phi_L \geq 0$ and $\lambda r^{\gamma_H - \gamma_L} n^{\phi_H - \phi_L} \geq 1$, both $\Delta \log w_{Hj}^A$ and $\Delta \log w_{Lj}^A$ are weakly increasing in r_j and n_j . The increase in r_j is strict if $\gamma_H > 0$, $\gamma_L > 0$, or the displayed dominance inequality is strict; the increase in n_j is strict if $\phi_H > 0$ or $\phi_L > 0$.

Proof. With agglomeration, output is $Y_{jt} = (A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho)^{1/\rho}$, and competitive factor markets

imply

$$w_{Hjt} = A_{Hjt} H_j^{\rho-1} (A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho)^{\frac{1}{\rho}-1}, \quad w_{Ljt} = A_{Ljt} L_j^{\rho-1} (A_{Hjt} H_j^\rho + A_{Ljt} L_j^\rho)^{\frac{1}{\rho}-1}.$$

Let $r_j \equiv H_j/L_j$ and normalize $L_j = 1$ (so $H_j = r_j$). At $t = 0$, $A_{Hj0} = A_{Lj0} = 1$, so

$$w_{Hj0} = r_j^{\rho-1} (1 + r_j^\rho)^{\frac{1}{\rho}-1}, \quad w_{Lj0} = (1 + r_j^\rho)^{\frac{1}{\rho}-1}.$$

At $t = 1$, $A_{Hj1} = \lambda r_j^{\gamma_H} n_j^{\phi_H}$ and $A_{Lj1} = r_j^{\gamma_L} n_j^{\phi_L}$, hence

$$w_{Hj1}^A = \lambda r_j^{\gamma_H} n_j^{\phi_H} r_j^{\rho-1} (\lambda r_j^{\gamma_H+\rho} n_j^{\phi_H} + r_j^{\gamma_L} n_j^{\phi_L})^{\frac{1}{\rho}-1},$$

$$w_{Lj1}^A = r_j^{\gamma_L} n_j^{\phi_L} (\lambda r_j^{\gamma_H+\rho} n_j^{\phi_H} + r_j^{\gamma_L} n_j^{\phi_L})^{\frac{1}{\rho}-1}.$$

Taking log differences yields the stated formulas.

Monotonicity in r . Differentiating and collecting $\theta'(r)$ terms gives

$$\frac{\partial F^A}{\partial r}(r, n) = \lambda \gamma_H r^{\gamma_H-1} n^{\phi_H} \theta(r) + \gamma_L r^{\gamma_L-1} n^{\phi_L} (1 - \theta(r)) + \theta'(r) (\lambda r^{\gamma_H} n^{\phi_H} - r^{\gamma_L} n^{\phi_L}).$$

Thus, if $\gamma_H, \gamma_L \geq 0$ and $\lambda r^{\gamma_H-\gamma_L} n^{\phi_H-\phi_L} \geq 1$, then $\partial F^A/\partial r \geq 0$ and $\log F^A$ is weakly increasing in r .

Monotonicity in n . Holding r fixed,

$$\frac{\partial F^A}{\partial n}(r, n) = \lambda \phi_H r^{\gamma_H} n^{\phi_H-1} \theta(r) + \phi_L r^{\gamma_L} n^{\phi_L-1} (1 - \theta(r)).$$

Hence, if $\phi_H, \phi_L \geq 0$, then $\partial F^A/\partial n \geq 0$ and $\log F^A$ is weakly increasing in n .

Combining these facts with the explicit wage-growth formulas above implies that $\Delta \log w_{Hj}^A$ and $\Delta \log w_{Lj}^A$ are weakly increasing in r_j and n_j on the stated domain. $\square$

Remark: initial average wages and β -patterns. In the baseline normalization $A_{Hj0} = A_{Lj0} = 1$, and setting $L_j = 1$ so that $H_j = r_j$, output is $Y_{j0} = (1 + r_j^\rho)^{1/\rho}$. The average wage (output per worker) is

$$w_{j0} = \frac{Y_{j0}}{1 + r_j} = \frac{(1 + r_j^\rho)^{1/\rho}}{1 + r_j}, \quad 0 < \rho < 1.$$

A direct derivative gives

$$\frac{d}{dr_j} \log w_{j0} = \frac{r_j^{\rho-1} - 1}{(1 + r_j^\rho)(1 + r_j)},$$

so w_{j0} is increasing in r_j for $r_j < 1$ and decreasing for $r_j > 1$. Consequently, the sign of a cross-sectional β -regression of average wage growth $\Delta \log w_j$ on initial average wages $\log w_{j0}$ is not pinned down mechanically by r_j alone; it depends on the joint distribution of (r_j, n_j) and other city fundamentals. Under the stated parameter and support conditions, the skill-premium change and group-specific wage growth rise with initial skill abundance, as

shown in Propositions 1–2 and 4–5. Proposition 6 then extends this logic to average wage growth.

Economic intuition. In addition to skill composition, overall size n_j now magnifies productivity when $\phi_H > \phi_L$. This introduces a size-based amplification consistent with empirical agglomeration effects.

Proposition 6 (Average wage growth with agglomeration). *With $\mu_H(r) = r/(1+r)$, $\mu_L(r) = 1/(1+r)$,*

$$\Delta \log w_j^A = \mu_H(r_j) \Delta \log w_{Hj}^A + \mu_L(r_j) \Delta \log w_{Lj}^A.$$

Under the conditions of Proposition 5, average wage growth is weakly increasing in r_j and n_j . The increase in r_j is strict if the technology-dominance inequality is strict or either group-specific r_j -derivative is strictly positive. The increase in n_j is strict if $\phi_H > 0$ or $\phi_L > 0$.

Proof. By definition,

$$\Delta \log w_j^A = \mu_H(r_j) \Delta \log w_{Hj}^A + \mu_L(r_j) \Delta \log w_{Lj}^A, \quad \mu_H(r) = \frac{r}{1+r}, \quad \mu_L(r) = \frac{1}{1+r}.$$

Monotonicity in r . Differentiate:

$$\frac{d}{dr} \Delta \log w_j^A = \mu'_H(r_j) (\Delta \log w_{Hj}^A - \Delta \log w_{Lj}^A) + \mu_H(r_j) \frac{d}{dr} \Delta \log w_{Hj}^A + \mu_L(r_j) \frac{d}{dr} \Delta \log w_{Lj}^A.$$

Since $\mu'_H(r) = 1/(1+r)^2 > 0$ and

$$\Delta \log w_{Hj}^A - \Delta \log w_{Lj}^A = \log \lambda + (\gamma_H - \gamma_L) \log r_j + (\phi_H - \phi_L) \log n_j,$$

the first term is nonnegative on any support where $\lambda r_j^{\gamma_H - \gamma_L} n_j^{\phi_H - \phi_L} \geq 1$. Under the conditions of Proposition 5 that ensure $\frac{d}{dr} \Delta \log w_{Hj}^A \geq 0$ and $\frac{d}{dr} \Delta \log w_{Lj}^A \geq 0$, it follows that $\frac{d}{dr} \Delta \log w_j^A \geq 0$.

Monotonicity in n . Because $\mu_H(r_j)$ and $\mu_L(r_j)$ do not depend on n_j ,

$$\frac{\partial}{\partial n_j} \Delta \log w_j^A = \mu_H(r_j) \frac{\partial}{\partial n_j} \Delta \log w_{Hj}^A + \mu_L(r_j) \frac{\partial}{\partial n_j} \Delta \log w_{Lj}^A,$$

so $\partial \Delta \log w_j^A / \partial n_j \geq 0$ whenever both group-specific wage-growth terms are weakly increasing in n_j (as ensured, for example, by $\phi_H, \phi_L \geq 0$ in Proposition 5). $\square$

Empirical mapping. Proposition 4 decomposes $\Delta \pi_j^A$ into a national SBTC component $\log \lambda$, a skill-composition amplifier $(\gamma_H - \gamma_L) \log r_j$, and a size amplifier $(\phi_H - \phi_L) \log n_j$. These components motivate the empirical comparisons by skill composition and city size; the full model evaluates their implications for regional convergence.

E.3 Time-Varying Skill Bias

The pace of SBTC may vary over time, as discussed by [Acemoglu and Autor \(2011\)](#). Let $\lambda_t > 0$ denote the skill-bias factor in the post-benchmark scenario indexed by $t \geq 1$. Its growth may slow while its level continues to rise. The notation λ_t is distinct from the quantitative scaling term Λ_{kt} in the main text.

Setup. Keep H_j, L_j , the agglomeration elasticities, and the common initial benchmark $A_{Hj0} = A_{Lj0} = 1$ fixed. Each scenario introduces the composition and size shifters together with λ_t . Throughout this subsection, $\Delta\pi_{j,t} \equiv \pi_{j,t} - \pi_{j,0}$ and $\Delta \log w_{kj,t} \equiv \log w_{kj,t} - \log w_{kj,0}$ compare scenario t with date zero. These comparisons isolate the persistence of cross-city differences relative to the same benchmark.

Proposition 7 (Cross-sectional premium changes with time-varying SBTC). *For any $t \geq 1$, the change in the log skill premium in city j relative to date zero is*

$$\Delta\pi_{j,t} = \log \lambda_t + (\gamma_H - \gamma_L) \log r_j + (\phi_H - \phi_L) \log n_j.$$

Hence the cross-sectional gradients

$$\frac{\partial \Delta\pi_{j,t}}{\partial \log r_j} = \gamma_H - \gamma_L \quad \text{and} \quad \frac{\partial \Delta\pi_{j,t}}{\partial \log n_j} = \phi_H - \phi_L$$

are independent of λ_t . A lower level of λ_t shifts premium changes downward uniformly and leaves these cross-city gradients unchanged. Slower growth in λ_t reduces the growth of the common component; it need not reduce its level.

Proof. In scenario t , write $A_{Hj1,t} = \lambda_t r_j^{\gamma_H} n_j^{\phi_H}$ and $A_{Lj1,t} = r_j^{\gamma_L} n_j^{\phi_L}$. By the same steps as in Propositions 1 and 4, the change in the log skill premium relative to date zero is

$$\Delta\pi_{j,t} = \log\left(\frac{A_{Hj1,t}}{A_{Lj1,t}}\right) + (\rho - 1) \log r_j - \left[\log\left(\frac{A_{Hj0}}{A_{Lj0}}\right) + (\rho - 1) \log r_j \right] = \log \lambda_t + (\gamma_H - \gamma_L) \log r_j + (\phi_H - \phi_L) \log n_j,$$

because $A_{Hj0} = A_{Lj0} = 1$. Differentiation w.r.t. $\log r_j$ and $\log n_j$ yields the stated slopes, which do not involve λ_t . $\square$

Proposition 8 (Wage-growth slopes with time-varying SBTC). *Let $\theta(r) \equiv r^\rho / (1 + r^\rho)$ and define*

$$F_t(r, n) := \lambda_t r^{\gamma_H} n^{\phi_H} \theta(r) + r^{\gamma_L} n^{\phi_L} [1 - \theta(r)].$$

Then for any $t \geq 1$,

$$\begin{aligned}\Delta \log w_{Hj,t} &= \log \lambda_t + \gamma_H \log r_j + \phi_H \log n_j + \frac{1-\rho}{\rho} \log F_t(r_j, n_j), \\ \Delta \log w_{Lj,t} &= \gamma_L \log r_j + \phi_L \log n_j + \frac{1-\rho}{\rho} \log F_t(r_j, n_j).\end{aligned}$$

Suppose $\gamma_H > \gamma_L \geq 0$, $\phi_H, \phi_L \geq 0$, and, for every $t \geq t_0$ and every (r, n) in the relevant support,

$$\lambda_t r^{\gamma_H - \gamma_L} n^{\phi_H - \phi_L} \geq 1.$$

Then for all $t \geq t_0$ and all (r_j, n_j) in that support,

$$\frac{\partial \Delta \log w_{Hj,t}}{\partial \log r_j} > 0, \quad \frac{\partial \Delta \log w_{Lj,t}}{\partial \log r_j} > 0,$$

and

$$\frac{\partial \Delta \log w_{Hj,t}}{\partial \log n_j} \geq 0, \quad \frac{\partial \Delta \log w_{Lj,t}}{\partial \log n_j} \geq 0.$$

The inequalities with respect to n_j are strict if $\phi_H > 0$ or $\phi_L > 0$.

Proof. The wage-growth formulas follow exactly as in Proposition 5, replacing λ by λ_t :

$$\Delta \log w_{Hj,t} = \log \lambda_t + \gamma_H \log r_j + \phi_H \log n_j + \frac{1-\rho}{\rho} \log F_t(r_j, n_j),$$

$$\Delta \log w_{Lj,t} = \gamma_L \log r_j + \phi_L \log n_j + \frac{1-\rho}{\rho} \log F_t(r_j, n_j),$$

where

$$F_t(r, n) = \lambda_t r^{\gamma_H} n^{\phi_H} \theta(r) + r^{\gamma_L} n^{\phi_L} (1 - \theta(r)), \quad \theta(r) = \frac{r^\rho}{1 + r^\rho}.$$

Derivatives with respect to r . As before,

$$\frac{\partial F_t}{\partial r}(r, n) = \lambda_t \gamma_H r^{\gamma_H - 1} n^{\phi_H} \theta(r) + \gamma_L r^{\gamma_L - 1} n^{\phi_L} (1 - \theta(r)) + \theta'(r) (\lambda_t r^{\gamma_H} n^{\phi_H} - r^{\gamma_L} n^{\phi_L}),$$

with $\theta'(r) > 0$. Hence, if $\gamma_H, \gamma_L \geq 0$ and (for all (r, n) in the empirical support and all $t \geq t_0$) $\lambda_t r^{\gamma_H - \gamma_L} n^{\phi_H - \phi_L} \geq 1$, then $\partial F_t / \partial r \geq 0$ on that support and $\log F_t$ is weakly increasing in r . Therefore,

$$\frac{\partial}{\partial \log r} \Delta \log w_{H,t} = \gamma_H + \frac{1-\rho}{\rho} r \frac{\partial F_t / \partial r}{F_t} \geq \gamma_H > 0, \quad \frac{\partial}{\partial \log r} \Delta \log w_{L,t} = \gamma_L + \frac{1-\rho}{\rho} r \frac{\partial F_t / \partial r}{F_t} \geq \gamma_L \geq 0,$$

with strict positivity for the less skilled derivative when either $\gamma_L > 0$ or $\partial F_t / \partial r > 0$.

Derivatives with respect to n . Holding r fixed,

$$\frac{\partial F_t}{\partial n}(r, n) = \lambda_t \phi_H r^{\gamma_H} n^{\phi_H - 1} \theta(r) + \phi_L r^{\gamma_L} n^{\phi_L - 1} (1 - \theta(r)),$$

so if $\phi_H, \phi_L \geq 0$, then $\partial F_t / \partial n \geq 0$ and $\log F_t$ is weakly increasing in n . It follows that

$$\frac{\partial}{\partial \log n} \Delta \log w_{H,t} \geq \phi_H, \quad \frac{\partial}{\partial \log n} \Delta \log w_{L,t} \geq \phi_L,$$

again with strict inequalities when the relevant terms are strictly positive. $\square$

Implication. Cross-city gradients in premium changes relative to date zero are invariant to λ_t . Under the stated support conditions, wage changes relative to that benchmark remain increasing in skill abundance and weakly increasing in city size, although their gradients can vary with λ_t . The fixed populations allow these composition and size channels to be characterized directly; migration and endogenous sorting enter the full quantitative model.

Connection to the Quantitative Model. The illustration separates national skill bias from composition and city-size contributions to wage changes, holding populations fixed and using a common initial-productivity benchmark. These comparative statics clarify the productivity channels in equation (18). The full model in Section 3 combines these channels with endogenous migration, sorting, diffusion, housing, and trade; its counterfactuals evaluate their contributions to skill-specific convergence.

References

- Acemoglu, Daron, and David Autor.** 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” , ed. O. Ashenfelter and D. Card Vol. 4 of *Handbook of Labor Economics*, Chapter 12, 1043–1171. Elsevier.
- Autor, David H., David Dorn, and Gordon H. Hanson.** 2015. “Untangling Trade and Technology: Evidence from Local Labour Markets.” *The Economic Journal*, 125(584): 621–646.
- Bajona, Claustre, and Timothy Kehoe.** 2010. “Trade, Growth, and Convergence in a Dynamic Heckscher-Ohlin Model.” *Review of Economic Dynamics*, 13(3): 487–513.
- Barro, Robert J., and Xavier Sala-i Martin.** 1997. “Technological Diffusion, Convergence, and Growth.” *Journal of Economic Growth*, 2(1): 1–26.
- Beaudry, Paul, Mark Doms, and Ethan Lewis.** 2010. “Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas.” *Journal of Political Economy*, 118(5): 988 – 1036.
- Benevieri, Pierluigi, Massimo Furi, Maria Patrizia Pera, and Marco Spadini.** 2023. “An Introduction to Topological Degree in Euclidean Spaces.”
- Caliendo, Lorenzo.** 2011. “On the Dynamics of the Heckscher-Ohlin Theory.” Working paper.
- Caselli, Francesco, and Wilbur John Coleman.** 2001. “The U.S. Structural Transformation and Regional Convergence: A Reinterpretation.” *Journal of Political Economy*, 109(3): 584–616.
- Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg.** 2018. “The Geography of Development.” *Journal of Political Economy*, 126(3): 903–983.
- Diamond, Rebecca.** 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review*, 106(3): 479–524.
- Eckert, Fabian, and Michael Peters.** 2022. “Spatial Structural Change.” National Bureau of Economic Research NBER Working Paper 30489.

- Eckert, Fabian, Sharat Ganapati, and Conor Walsh.** 2022. “Urban-Biased Growth: A Macroeconomic Analysis.” National Bureau of Economic Research NBER Working Paper 30515.
- Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny.** 2014. “Spatial Sorting.” *Journal of Political Economy*, 122(3): 554–620.
- Faberman, R. Jason, and Matthew Freedman.** 2016. “The urban density premium across establishments.” *Journal of Urban Economics*, 93(C): 71–84.
- Feenstra, Robert, and Gordon Hanson.** 1999. “The Impact of Outsourcing and High-Technology Capital on Wages: Estimates For the United States, 1979–1990.” *The Quarterly Journal of Economics*, 114(3): 907–940.
- Ganong, Peter, and Daniel Shoag.** 2017. “Why Has Regional Income Convergence in the U.S. Declined?” *Journal of Urban Economics*, 102: 76–90.
- Gyourko, Joseph, Albert Saiz, and Anita Summers.** 2008. “A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index.” *Urban studies*, 45(3): 693–729.
- Gyourko, Joseph, Christopher Mayer, and Todd Sinai.** 2013. “Superstar cities.” *American Economic Journal: Economic Policy*, 5(4): 167–199.
- Holmes, Thomas J.** 1998. “The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders.” *Journal of Political Economy*, 106(4): 667–705.
- Hsieh, Chang-Tai, and Enrico Moretti.** 2019. “Housing Constraints and Spatial Misallocation.” *American Economic Journal: Macroeconomics*, 11(2): 1–39.
- Kennan, John, and James R Walker.** 2011. “The effect of expected income on individual migration decisions.” *Econometrica*, 79(1): 211–251.
- Kleinman, Benny, Ernest Liu, and Stephen J. Redding.** 2023. “Dynamic Spatial General Equilibrium.” *Econometrica*, 91(2): 385–424.
- Notowidigdo, Matthew J.** 2020. “The Incidence of Local Labor Demand Shocks.” *Journal of Labor Economics*, 38(3): 687–725.
- Rubinton, Hannah.** 2025. “The Geography of Business Dynamism and Skill-Biased Technical Change.” *The Review of Economic Studies*, rda063.

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

OA.1 Online Data Appendix

The paper studies the difference in wage-convergence trends shown in Figure 2 and Section 2.2: β^H shifts from negative pre-1980 to approximately zero post-1980, while β^L remains negative. I examine how this pattern varies with metropolitan coverage, skill definitions, demographic composition, and the treatment of top-coded wages. Rolling-window estimates document the timing of the change in convergence. I first describe the data sources and then present these checks for aggregate wages and for each skill group.

OA.1.1 Data Description

My two main microdata sources are the IPUMS USA Census/ACS extracts and the March CPS. For the Census/ACS component, I use the 1% Census samples for 1940, 1950, and 1960; the 1% metropolitan sample for 1970; the 5% Census samples for 1980, 1990, and 2000; and the 2010 American Community Survey (ACS), all extracted from IPUMS. In 1950, some key income variables are available only for the sample-line person in each selected household, so the effective person-level coverage for those variables is smaller than 1% before any additional restrictions. I use the information on wages, education, age, race, ethnicity, rents, birthplace, migration, population, industries, occupation, MSA, and state. All of this information is also available in the ACS data for 2010. I collect the same information from the CPS. 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 March survey. The CPS is primarily used for the analysis of migration. My geographic unit of research is the MSA. An MSA is a “region consisting of a large urban core and surrounding communities with a high degree of economic and social integration with the urban core.” I also use two additional data sets: the Wharton Land Use Regulation Index (WLURI), aggregated to the MSA level by [Saiz \(2010\)](#), and the RTI measure developed by [Autor and Dorn \(2013\)](#). The latter uses occupational task measures from the 1977 Dictionary of Occupational Titles.³

OA.1.2 Robustness Checks for Fact 2

Before turning to these robustness tests, I restate the β -convergence specification used throughout the paper following (1). In most specifications, observations are weighted by the initial size of location j .

I run several robustness tests starting with the ones illustrated in Figure 1 and in Figure 2. I change the unit of analysis from cities to counties in Figure OA.1. In Figure OA.1, I plot the

³For a more detailed description of the RTI measure, please refer to [Autor and Dorn \(2013\)](#).

estimated convergence rates. The estimate uses a 10-year rolling period in plot A, while plot B uses a 20-year rolling period. The x-axis reports the window end year. The convergence rate is negative and statistically significant through end-year 1987 in plot A and through end-year 1997 in plot B. The first window that is not statistically different from zero starts in 1978 in both panels (1978–1988 in plot A and 1978–1998 in plot B). This finding aligns with [Higgins, Levy and Young \(2006\)](#), who find convergence between 1970 and 1990. However, departing from this prior work, I conduct an analysis in which the period is extended and find that the convergence across counties follows the same patterns as the convergence across cities and states.

As a second robustness check, I compute rolling 20- and 30-year regional convergence rates beginning in 1940, as illustrated in Figure [OA.2](#), and then decompose the estimates by skill group. Panels (c)–(d) and (e)–(f) report the results for highly skilled and less skilled workers, respectively. In the aggregate 20-year series, the coefficients are significantly negative in 1940–1960 and 1960–1980, statistically indistinguishable from zero in 1950–1970 and 1970–1990, and positive but insignificant in 1980–2000 and 1990–2010. The 30-year aggregate coefficients are significantly negative through 1960–1990 and insignificant thereafter. Thus, aggregate convergence weakens in windows ending around 1990. The positive aggregate coefficients for 1990–2010 are statistically indistinguishable from zero. The skill decomposition preserves the central Fact 2 pattern. Highly skilled wage convergence is significantly negative in the early 20-year windows, becomes positive but insignificant in 1980–2000 and 1990–2010, and is likewise positive but insignificant in the 1980–2010 30-year window. Less skilled wage convergence remains negative and statistically significant in every reported 20- and 30-year window.

As a third robustness check, I reproduce Figures [1](#) and [2](#) with compositionally adjusted wages. I estimate the demographic adjustment separately by year and skill group using age, sex, and birthplace-group controls, and then test whether the convergence rates change once wages are adjusted for these observable characteristics. As shown in Figure [OA.3](#), the convergence rates do not change substantially after this non-skill compositional adjustment.

To assess sensitivity to changes in metropolitan coverage, I re-estimate Fact 2 using a fixed sample of MSAs observed in 1940, 1960, and 1980. Panel A of Table [OA.2](#) shows that the main result is unchanged in this fixed MSA sample: before 1980, both groups converge ($\beta^L = -1.22$, $\beta^H = -1.30$, both significant), while after 1980 less skilled wages still converge ($\beta^L = -1.95$, significant) and highly skilled convergence is near zero and not statistically different from zero ($\beta^H = 0.15$). This check assesses sensitivity to changes in MSA coverage across decades. Additional checks in Panels B–E of Online Appendix Table [OA.2](#) confirm that the same divergence pattern is robust to compositionally adjusted wages, an alternative

skill binning (some college or more), and two explicit top-tail treatments of hourly wages (trim or winsorize the top 1% within year and skill). Together, these checks support a strong empirical basis for Fact 2.

OA.1.3 Convergence with SBTC Exposure Controls

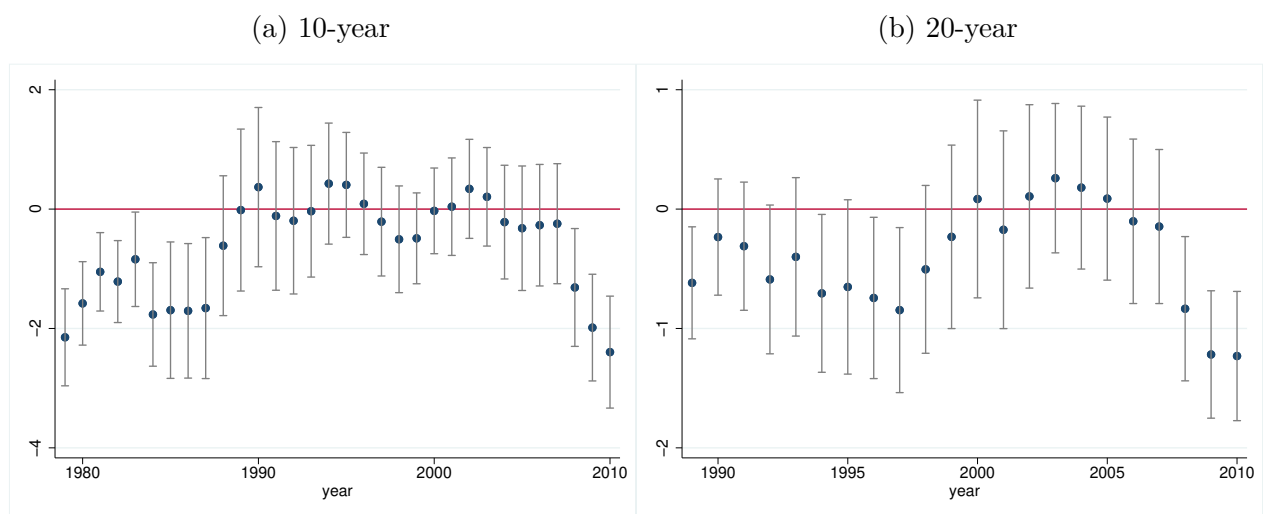
I also study whether local SBTC exposure helps explain the post-1980 decline in aggregate regional convergence. I regress annual average wage growth between 1980 and 2010 on the initial log wage in 1980 and then sequentially add contemporaneous and lagged SBTC exposure measures, $\Delta S_{Hj\tau}$ and $\Delta S_{Lj\tau}$. The outcome is wage growth for each MSA as a whole. Across these nested specifications, the point estimate of the β -convergence coefficient becomes progressively more negative, reaching -0.90 percentage points per year in column E of Table OA.3, but it remains statistically indistinguishable from zero at conventional levels. Several exposure coefficients—especially the 1970 less-skilled measure—are statistically significant.

Formally, column $m \in \{A, B, C, D, E\}$ of Table OA.3 estimates

$$g_j^{80-10} = \alpha + \beta \widetilde{\ln w_{j,1980}} + \sum_{\tau \in \mathcal{T}_m} (\gamma_{H,\tau} \Delta S_{Hj\tau} + \gamma_{L,\tau} \Delta S_{Lj\tau}) + u_j, \quad (\text{OA.1})$$

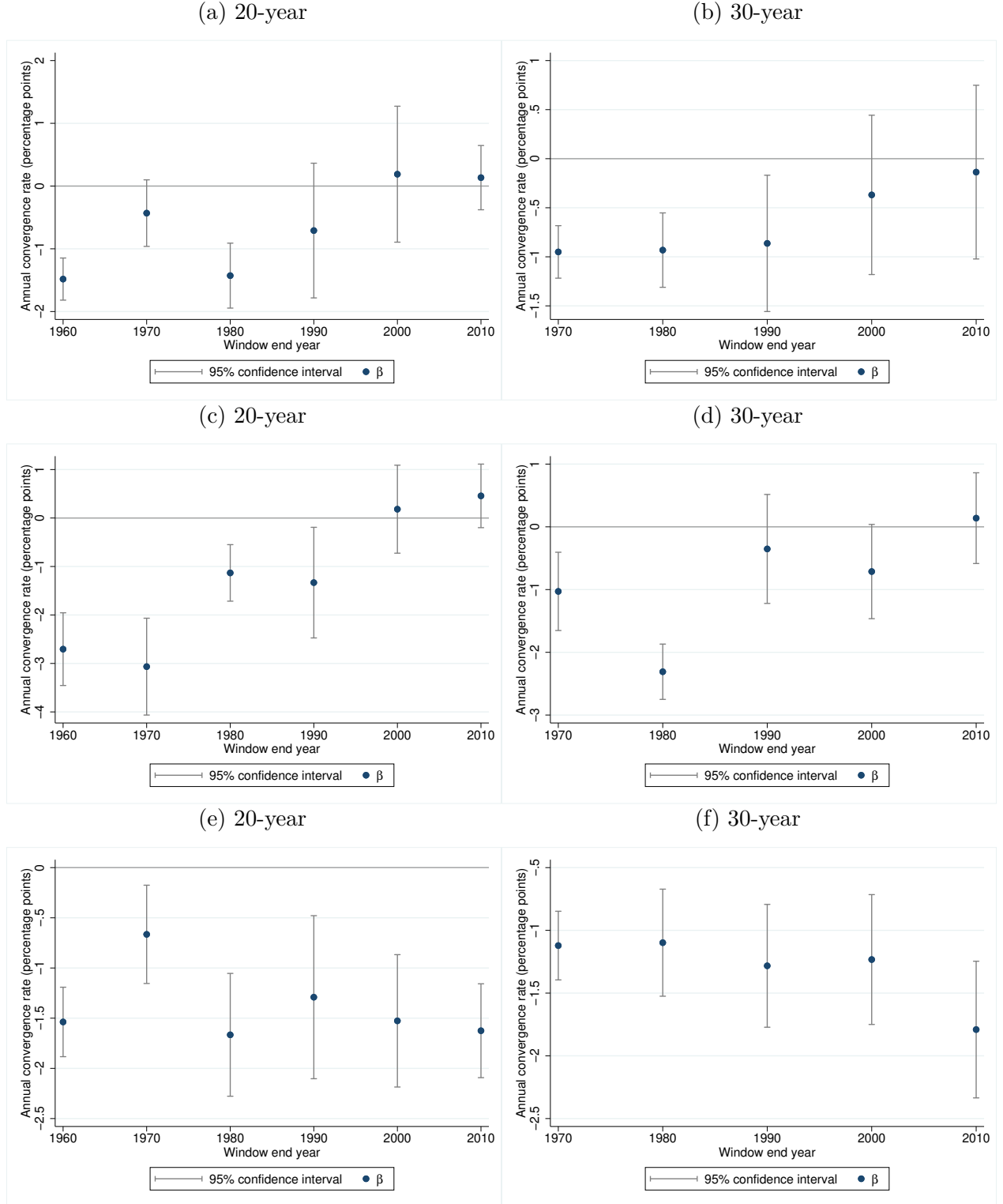
where $g_j^{80-10} \equiv 100 \times (\ln w_{j,2010} - \ln w_{j,1980})/30$ is annual average aggregate wage growth in percentage points, $\widetilde{\ln w_{j,1980}}$ is the demeaned initial log wage in 1980, and the set of included SBTC controls varies across columns as exposure measures from progressively earlier decades are added. All specifications are weighted by 1980 MSA population.

Figure OA.1: Convergence by County over Time



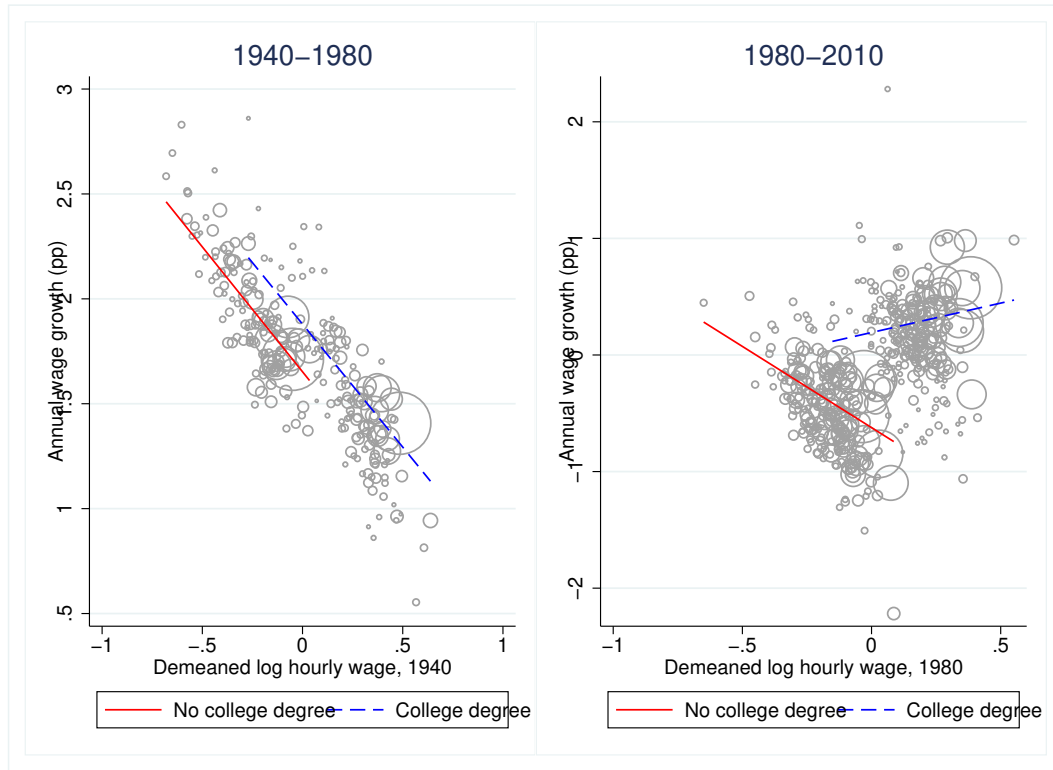
Note: Figure OA.1 shows estimates of β -convergence by county. Plot A shows the convergence rate at the county level for a 10-year rolling window that starts in 1969. Plot B shows the convergence rate at the county level for a 20-year rolling window that starts in 1969. Each point is indexed by end year t : plot A uses $[t - 10, t]$, and plot B uses $[t - 20, t]$. Data for this analysis are from the Bureau of Economic Analysis Regional Economics Accounts. In each estimate, counties are weighted by population. The y-axis reports the coefficient in percentage terms.

Figure OA.2: Convergence Rates over Time: Overall and by Skill Group



Note: Figure OA.2 shows the β estimates from regressions of annualized log wage growth on initial log wages using 20-year and 30-year rolling windows. In each estimate, cities are weighted by population. The y-axis reports the coefficient in percentage terms. Plots (a) and (b) report aggregate β ; plots (c)–(d) and (e)–(f) report β^H and β^L , respectively.

Figure OA.3: Compositionally Adjusted Regional Convergence



Note: Figure OA.3 shows two scatter plots of the log wages by MSA in the initial year against the annual average growth of wages in the final year. Wages are adjusted separately by year and skill group using age, sex, and birthplace-group controls before taking the MSA average. Specifically, on the left-hand side (right-hand side), I plot the demeaned log wages in 1940 (1980) by MSA against the annual average wage growth between 1940 (1980) and 1980 (2010). The line in each graph represents a weighted regression line from the bivariate regression.

Table OA.1: Convergence Rates: Robustness

	(1) Δw_{pw}^{40-80}	(2) Δw_{pw}^{80-10}	(3) Δw^{40-80}	(4) Δw^{80-10}
Panel A: Baseline				
$\ln(wage^{1940})$	-1.13*** (-10.24)		-1.49*** (-16.36)	
$\ln(wage^{1980})$		-0.14 (-0.30)		-1.05*** (-3.41)
Observations	127	239	127	239
Panel B: Fixed Pre-1980 MSA Sample				
$\ln(wage^{1940})$	-1.12*** (-10.02)		-1.47*** (-15.21)	
$\ln(wage^{1980})$		-0.21 (-0.36)		-1.34** (-2.58)
Observations	122	121	122	121
Panel C: Compositionally Adjusted Wages				
$\ln(wage^{1940})$	-1.05*** (-8.96)		-1.45*** (-14.70)	
$\ln(wage^{1980})$		0.17 (0.36)		-0.81*** (-2.64)
Observations	127	239	127	239
Panel D: Trim Top 1% Wage Tail				
$\ln(wage^{1940})$	-1.15*** (-10.88)		-1.43*** (-17.69)	
$\ln(wage^{1980})$		-0.39 (-0.99)		-1.07*** (-4.07)
Observations	127	239	127	239
Panel E: Winsorize Top 1% Wage Tail				
$\ln(wage^{1940})$	-1.15*** (-10.99)		-1.47*** (-17.60)	
$\ln(wage^{1980})$		-0.16 (-0.37)		-0.94*** (-3.26)
Observations	127	239	127	239

Note: Table OA.1 reports Fact 1 robustness checks for aggregate convergence. Panels A–E report, respectively, the baseline specification, fixed pre-1980 MSA sample, compositionally adjusted wages, top-tail trim, and top-tail winsorization. Columns (1)–(2) are population-weighted and columns (3)–(4) are unweighted. All reported β -convergence coefficients are annual rates measured in percentage points.

Table OA.2: Convergence by Skill: Robustness to Sample, Binning, and Top-Tail Treatment

	(1) Less Skilled 1940–1980	(2) Highly Skilled 1940–1980	(3) Less Skilled 1980–2010	(4) Highly Skilled 1980–2010
Panel A: Convergence by Skill in a Fixed Pre-1980 MSA Sample				
$\ln(wage^{1940})$	−1.22*** (−9.52)	−1.30*** (−6.39)		
$\ln(wage^{1980})$			−1.95*** (−5.52)	0.15 (0.30)
Observations	122	122	121	121
Panel B: Compositionally Adjusted Wages				
$\ln(wage^{1940})$	−1.191*** (−9.80)	−1.173*** (−5.63)		
$\ln(wage^{1980})$			−1.393*** (−3.77)	0.505 (1.22)
Observations	127	127	239	239
Panel C: Alternative Skill Binning (Some College+)				
$\ln(wage^{1940})$	−1.31*** (−10.41)	−1.01*** (−7.49)		
$\ln(wage^{1980})$			−1.99*** (−8.84)	0.42 (0.93)
Observations	127	127	239	239
Panel D: Trim Top 1% Wage Tail				
$\ln(wage^{1940})$	−1.20*** (−8.89)	−1.27*** (−6.63)		
$\ln(wage^{1980})$			−1.89*** (−7.41)	−0.22 (−0.88)
Observations	127	127	239	239
Panel E: Winsorize Top 1% Wage Tail				
$\ln(wage^{1940})$	−1.25*** (−9.76)	−1.25*** (−6.43)		
$\ln(wage^{1980})$			−1.81*** (−6.57)	0.21 (0.65)
Observations	127	127	239	239

Note: Table OA.2 reports Fact 2 robustness checks by skill group. Panel A uses the fixed pre-1980 MSA sample. Panel B reports compositionally adjusted wages. Panels C–E report alternative skill binning and top-tail trim/winsor treatments. All reported β -convergence coefficients are annual rates measured in percentage points.

Table OA.3: Convergence Rates and SBTC

	(1) A	(2) B	(3) C	(4) D	(5) E
$\ln(wage^{1980})$	-0.137 (-0.30)	-0.653 (-1.72)	-0.674 (-1.42)	-0.800 (-1.42)	-0.904 (-1.51)
ΔS_H^{1980}		1.068 (1.09)	1.088 (0.74)	1.036 (0.65)	1.274 (0.78)
ΔS_L^{1980}		3.847*** (3.35)	1.358 (0.58)	1.633 (0.67)	1.324 (0.54)
ΔS_H^{1970}			1.568* (2.15)	1.549 (1.92)	1.662 (1.92)
ΔS_L^{1970}			4.020* (2.30)	4.645* (2.27)	5.208* (2.24)
ΔS_H^{1950}				-0.165 (-0.53)	-0.254 (-0.78)
ΔS_L^{1950}				0.140 (0.20)	0.288 (0.34)

Note: Table OA.3 reports regressions of annual average aggregate wage growth between 1980 and 2010 on the demeaned initial log wage in 1980 and on SBTC exposure measures based on equations (31) and (30); see equation (OA.1). Column A reports the baseline convergence regression with no SBTC controls. Column B adds the 1980 and 2010 SBTC measures. Column C further adds the 1970 measures. Column D further adds the 1950 measures. Column E further adds the 1940 measures. To save space, the table displays the coefficients on the initial log wage and on the 1980, 1970, and 1950 SBTC measures; the coefficients on the 2010 SBTC measures in columns B–E and on the 1940 SBTC measures in column E are estimated but omitted from the printed table. All reported β -convergence coefficients are annual rates measured in percentage points.

References

- Autor, David H., and David Dorn.** 2013. “The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market.” *American Economic Review*, 103(5): 1553–1597.
- Higgins, Matthew, Daniel Levy, and Andrew Young.** 2006. “Growth and Convergence across the United States: Evidence from County-Level Data.” *The Review of Economics and Statistics*, 88(4): 671–681.
- Saiz, Albert.** 2010. “The Geographic Determinants of Housing Supply.” *Quarterly Journal of Economics*, 125(3): 1253–1296.